(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 October 2002 (03.10.2002)

PCT

(10) International Publication Number
WO 02/077180 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/08964

(22) International Filing Date: 20 March 2002 (20.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/815,925 22 March 2001 (22.03.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/815,925 (CON)
Filed on 22 March 2001 (22.03.2001)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvalle, CA 94086 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (CH). **ZHOU, Ping** [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). **GOODRICH, Ryle** [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **ZHANG, Jie** [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). **YANG, Yonghong** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WHERMAN, Tom** [US/US]; CCSR Mol Pharm 3210, 269 West Campus Drive, Stanford, California 94305 (US). **DRMANAC, Radoje, T.** [US/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P., C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(7)

WO 02/077180 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2. BACKGROUND OF THE INVENTION

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins), has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1 – 11 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1 – 11 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1 – 11. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1 – 11 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1 – 11. The sequence information can be a segment of any one of SEQ ID NO: 1 – 11 that uniquely identifies or represents the sequence information of SEQ ID NO: 1 – 11.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect

full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-11 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-11 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1–11; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1–11; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1–11. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1–11; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in

SEQ ID NO: 1-11; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such

antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a

compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms “a”, “an” and “the” include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms “biologically active” or “biological activity” refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise “immunologically active” or “immunological activity” refers to the capability of the natural,

recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5’-AGT-3’ binds to the complementary sequence 3’-TCA-5’. Complementarity between two single-stranded molecules may be “partial” such that only some of the nucleic acids bind or it may be “complete” such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term “embryonic stem cells (ES)” refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term “germ line stem cells (GSCs)” refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term “primordial germ cells (PGCs)” refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs is nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or “nucleic acid” or “polynucleotide” or “oligonucleotide” are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic

origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or “probe” or “primer” are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs: 1-11.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NOs: 1-11. The sequence information can be a segment of any one of SEQ ID NOs: 1-11 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-11. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing

the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine

residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is

expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature

sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1 – 11; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1 – 11; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1 – 11. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1 – 11; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic

variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 1-11. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include the entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1 – 11 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1 - 11 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1 - 11 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%,

88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1 - 11, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1 - 11, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NOs: 1 - 11 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NOs: 1 - 11, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotides which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may

also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-11, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g.,

plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 – 11 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NOs: 1 - 11 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the

host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1 - 11, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1 - 11 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1 - 11 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1 - 11, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or

genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1 - 11). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved

in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the

advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association

with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et

al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of

inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No.

PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 1-11 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NOs: 1 – 11 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NOs: 1 – 11 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 1-11 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 1-11 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 1-11.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in

a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology.* Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 1-11.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.*, as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP- HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to

immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are

fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into

such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo, in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in

the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the

targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human

mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can

be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as

chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds.

Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which

would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and

identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I.

Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other

tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine,

kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as

candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of

human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor

specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine

169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer,

larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate,

Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g.. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between

polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be

assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic

(granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal

muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be

used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The

polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity

and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti- inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor,

thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a

therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels,

syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in

unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may

be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively

antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described

above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the

protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda

chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 1-11, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish

the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention

further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques,

such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the

hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and

light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at

least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host

have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin

light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been

produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic

protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present

invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NOs: 1 - 11 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NOs: 1 - 11 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl.

Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One

skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present

invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NOs: 1 – 11, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient

to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA

or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivative· which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NOs: 1 - 11. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NOs: 1 - 11 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.,* (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via a phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm$_7$), is then added to a final concentration of 10 mM 1-MeIm$_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm$_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally

equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5. EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

5.2 EXAMPLE 2

Novel Nucleic Acids

The novel nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri and UniGene) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above

databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST, gb pri, UniGene, Genpept) and the amino acid version of Genseq. Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1- 11.

Table 1 shows the various tissue sources of SEQ ID NO: 1-11.

The nearest neighbor results for polypeptides encoded by SEQ ID NO: 1-11 were obtained by a BLASTP (BLAST 1.2.3-Paracell (2001-11-20)) search against Genpept, Genseq and SwissProt databases using BLAST algorithm. The nearest neighbor result showed the closest homologue with functional annotation for SEQ ID NO: 1-11. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologues with identifiable functions for SEQ ID NO: 1-11 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol. Vol. 6 pp. 219-235 (1999) herein incorporated by reference), polypeptides encoded by SEQ ID NO: 1-11 were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) polypeptides encoded by SEQ ID NO: 1-11 were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the product of all the e-value of similar domains found, the Pfam score for the identified domain within the sequence, number of similar domains found, and the position of the domain in the SEQ ID NO: being interrogated.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ

ID NO: 1-11. Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al., (Nucl. Acids Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA) which is an automated sequence and structure searching procedure (http://www.msi.com/), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows, "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (http://www.rcsb.org/PDB/); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (U.S. Patent No. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:12502-13597. The verify score produced by GeneAtlas™ normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score - 1/2 high score)/(1/2 high score)

The PMF score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potential (MFP). As given in Table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0 with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based on all the information in Table 5 taken in totality.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determined from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for

identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol 10, no. 1, pp. 1-16 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al., as reference, were obtained for the polypeptide sequences. Table 6 shows the position of the last amino acid of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

Table 7 correlates each of SEQ ID NO: 1-11 to a specific chromosomal location.

Table 8 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-11, novel polypeptide sequences SEQ ID NO: 1-11, and their corresponding priority nucleotide sequences in the priority application USSN 09/815,925, herein incorporated by reference in its entirety.

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
adult brain	BioChain	ABR012	3
adult brain	Clontech	ABR001	3
adult brain	Clontech	ABR008	3 6 10
adult brain	GIBCO	ABD003	3
adult heart	GIBCO	AHR001	3 5
adult kidney	GIBCO	AKD001	3 7 9
adult kidney	Invitrogen	AKT002	6 9
adult lung	GIBCO	ALG001	6 9
adult spleen	Clontech	SPLc01	9
bladder	Invitrogen	BLD001	1 5
bone marrow	Clontech	BMD001	9
bone marrow	GF	BMD002	1 3
cervix	BioChain	CVX001	7 9
colon	Invitrogen	CLN001	2 9
endothelial cells	Strategene	EDT001	3 6-7 9-10
fetal brain	Clontech	FBR006	3 10-11
fetal brain	GIBCO	HFB001	3
fetal heart	Invitrogen	FHR001	9
fetal kidney	Clontech	FKD002	7
fetal liver-spleen	Soares	FLS001	3 5 9
fetal liver-spleen	Soares	FLS002	3-6 9
fetal lung	Invitrogen	FLG003	1-2 6-7
fetal muscle	Invitrogen	FMS001	3 8 11
fetal muscle	Invitrogen	FMS002	6
fetal skin	Invitrogen	FSK001	1 6
infant brain	Soares	IB2002	1 3 9
leukocytes	GIBCO	LUC001	9
lymphocyte	ATCC	LPC001	9
mix	B/I/C	SUP008	9
mix	B/I/C	SUP009	3
mixed		CGd011	4
mixed		CGd012	2 4 6 8 11
neuron	Strategene	NTD001	9
ovary	Invitrogen	AOV001	3 7 9
prostate	Clontech	PRT001	7
rectum	Invitrogen	REC001	3
salivary gland	Clontech	SAL001	2
skeletal muscle	Clonetech	SKMS03	8
skeletal muscle	Clontech	SKM001	3 8 11
small intestine	Clontech	SIN001	8
thymus	Clontech	THMc02	1
trachea	Clontech	TRC001	3
umbilical cord	BioChain	FUC001	3 6
uterus	Clontech	UTR001	1
young liver	GIBCO	ALV001	9

TABLE 2

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
12	gi11275980	Homo sapiens	NOTCH 1 (N1) mRNA, complete cds.	14525	99
12	gi338675	Homo sapiens	Human TAN-1 mRNA (homologue of Drosophila Notch gene), 5' end.	13918	99
12	gi3123675	Rattus rattus	rat notch protein	13122	89
13	gi2370133	Homo sapiens	partial MUC5B gene, exon 30-48.	3553	98
13	gi2290534	Homo sapiens	Human sublingual gland mucin (MUC5B) mRNA, partial cds.	3494	99
13	gi12843546	Mus musculus	putative	2265	62
14	AAB47140	Homo sapiens	INCY- CDIFF-21, Incyte ID No. 1360522CD1.	959	100
14	AAB37424	Homo sapiens	ROSE/ Human secreted protein BLAST search protein SEQ ID NO: 134.	959	100
14	AAB36111	Homo sapiens	CURA- Human MTC50.	959	100
15	gi4377339	Chlamydophila pneumoniae CWL029	Methionine Aminopeptidase	80	34
15	gi8979382	Chlamydophila pneumoniae J138	methionine aminopeptidase	80	34
15	gi9800317	rat cytomegalovirus Maastricht	pR102	78	29
16	gi15081392	Homo sapiens	NAC1 protein mRNA, complete cds.	2742	100
16	gi2384732	Rattus norvegicus	NAC-1 protein	2328	87
16	gi12849997	Mus musculus	putative	2325	87
17	AAH46863_aa1	Homo sapiens	MILL- Human 23553 sulfatase polypeptide encoding cDNA.	3137	100
17	AAB85483	Homo sapiens	MILL- Human 23553 sulfatase polypeptide.	3137	100
17	gi15394725	Homo sapiens	unnamed protein product	3137	100
18	gi12838404	Mus musculus	putative	1581	90
18	AAU18309	Homo sapiens	HUMA- Human endocrine polypeptide SEQ ID No 264.	1507	99
18	AAU19635	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 285.	679	97
19	gi12844321	Mus musculus	putative	2979	92
19	gi16552019	Homo sapiens	cDNA FLJ32015 fis, clone NTONG1000052, weakly similar to Rattus norvegicus mRNA for Kelch related protein 1.	2284	99
19	gi16306813	Homo sapiens	clone MGC:1367 IMAGE:2959774, mRNA, complete cds.	1715	50
20	AAB95294	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17519.	2432	100

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
20	gi10434931	Homo sapiens	cDNA FLJ13081 fis, clone NT2RP3002033.	2432	100
20	AAB85773	Homo sapiens	INCY- Human drug metabolizing enzyme (ID No. 19323521D1).	2405	99
21	gi182925	Homo sapiens	Human gamma amino butyric acid (GABAA) receptor beta-3 subunit mRNA, complete cds.	2348	99
21	gi14714965	Homo sapiens	gamma-aminobutyric acid (GABA) A receptor, beta 3, clone MGC:9051 IMAGE:3871111, mRNA, complete cds.	2341	99
21	gi755159	Mus musculus	GABA-benzodiazepine receptor beta-3 subunit	2307	97
22	AAM95118	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3776.	650	84
22	gi12838772	Mus musculus	putative	478	55
22	ABB12186	Homo sapiens	HYSE- Human AQ homologue, SEQ ID NO:2556.	352	64

TABLE 3

SEQ ID NO:	Database entry ID	Description	Results*
12	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 8.500e-11 1959-1972
12	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 7.164e-10 1066-1087 PR00764F 16.89 7.164e-10 1190-1211
12	PF00023	Ank repeat proteins.	PF00023B 14.20 6.000e-10 1962-1972 PF00023A 16.03 7.375e-10 2066-2082
12	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 1.342e-09 2404-2417
12	BL00612	Osteonectin domain proteins.	BL00612B 11.35 1.379e-09 1310-1343
12	BL01185	C-terminal cystine knot proteins.	BL01185B 21.14 2.636e-11 1140-1189 BL01185B 21.14 8.948e-11 747-796 BL01185B 21.14 5.667e-10 940-989 BL01185B 21.14 3.118e-09 291-340
12	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 3.475e-09 761-789
12	PR00680	P-TYPE TREFOIL DOMAIN SIGNATURE	PR00680B 10.34 7.682e-09 1506-1519
12	PR00009	TYPE I EGF SIGNATURE	PR00009C 14.11 8.059e-09 659-671
12	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 9.214e-10 2033-2088 PF00791B 28.49 8.514e-09 2000-2055
12	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 5.154e-16 1245-1261 BL01187B 12.04 8.650e-14 429-445 BL01187B 12.04 9.100e-14 885-901 BL01187B 12.04 2.565e-13 999-1015 BL01187B 12.04 5.304e-13 312-328 BL01187B 12.04 2.667e-12 1161-1177 BL01187B 12.04 5.667e-12 693-709 BL01187B 12.04 7.333e-12 730-746 BL01187B 12.04 7.667e-12 655-671 BL01187B 12.04 9.000e-12 961-977 BL01187B 12.04 1.600e-11 467-483 BL01187B 12.04 4.300e-11 1037-1053 BL01187B 12.04 4.900e-11 543-559 BL01187A 9.98 2.714e-10 944-956 BL01187B 12.04 2.800e-10 580-596 BL01187B 12.04 3.314e-10 505-521 BL01187A 9.98 4.857e-10 1182-1194 BL01187B 12.04 6.143e-10 350-366 BL01187B 12.04 7.943e-10 272-288 BL01187B 12.04 8.457e-10 1123-1139 BL01187A 9.98 8.714e-10 450-462 BL01187B 12.04 9.486e-10 155-171 BL01187A 9.98 2.500e-09 1266-1278 BL01187B 12.04 3.700e-09 806-822 BL01187B 12.04 3.925e-09 1199-1215 BL01187A 9.98 4.000e-09 868-880 BL01187B 12.04 6.400e-09 195-211 BL01187A 9.98 8.500e-09 488-500 BL01187A 9.98 8.500e-09 564-576 BL01187B 12.04 8.875e-09 1075-1091

*Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence

SEQ ID NO:	Database entry ID	Description	Results*
12	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 7.750e-12 684-694 DM00060 6.92 1.750e-11 914-924 DM00060 6.92 7.250e-11 759-769 DM00060 6.92 9.750e-11 146-156 DM00060 6.92 6.850e-10 108-118 DM00060 6.92 4.780e-09 224-234 DM00060 6.92 4.960e-09 303-313 DM00060 6.92 9.100e-09 1314-1324
12	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919A 11.53 1.818e-09 339-351 PD00919A 11.53 5.227e-09 1150-1162 PD00919A 11.53 6.045e-09 144-156 PD00919A 11.53 6.318e-09 532-544 PD00919A 11.53 8.500e-09 912-924 PD00919A 11.53 9.182e-09 950-962
12	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 9.852e-11 1287-1306 PR00011B 13.08 5.174e-10 236-255 PR00011B 13.08 3.466e-09 1287-1306 PR00011D 14.03 7.261e-09 545-564 PR00011D 14.03 7.391e-09 236-255 PR00011A 14.06 8.151e-09 236-255 PR00011D 14.03 8.174e-09 582-601 PR00011B 13.08 9.260e-09 1125-1144
12	BL00279	Membrane attack complex components / perforin proteins.	BL00279E 37.11 5.385e-11 1049-1097 BL00279E 37.11 7.923e-11 1173-1221 BL00279E 37.11 3.711e-10 129-177 BL00279E 37.11 5.012e-10 742-790 BL00279E 37.11 8.048e-10 441-489 BL00279E 37.11 2.047e-09 1011-1059 BL00279E 37.11 9.267e-09 897-945 BL00279E 37.11 9.581e-09 1338-1386
12	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 7.882e-13 1250-1261 PR00010C 11.16 7.750e-12 1004-1015 PR00010A 11.79 1.692e-11 791-803 PR00010A 11.79 3.077e-11 452-464 PR00010C 11.16 3.333e-11 966-977 PR00010A 11.79 2.452e-10 528-540 PR00010A 11.79 3.323e-10 678-690 PR00010C 11.16 3.455e-10 698-709 PR00010A 11.79 3.613e-10 946-958 PR00010A 11.79 7.677e-10 1184-1196 PR00010A 11.79 9.419e-10 490-502 PR00010C 11.16 1.214e-09 1331-1342 PR00010B 10.30 2.059e-09 690-698 PR00010C 11.16 2.929e-09 434-445 PR00010B 10.30 3.647e-09 996-1004 PR00010C 11.16 4.214e-09 160-171 PR00010C 11.16 4.429e-09 890-901 PR00010C 11.16 5.500e-09 83-94 PR00010C 11.16 6.143e-09 585-596 PR00010C 11.16 7.214e-09 1166-1177 PR00010C 11.16 7.429e-09 510-521 PR00010C 11.16 8.500e-09 355-366 PR00010C 11.16 9.571e-09 548-559

SEQ ID NO:	Database entry ID	Description	Results*
			PR00010C 11.16 9.786e-09 660-671
12	BL00022	EGF-like domain proteins.	BL00022B 7.54 3.250e-10 1254-1261 BL00022B 7.54 4.600e-09 1008-1015 BL00022A 7.48 9.000e-09 687-694 BL00022B 7.54 9.100e-09 127-134 BL00022B 7.54 9.100e-09 398-405 BL00022B 7.54 1.000e-08 1046-1053
13	BL01185	C-terminal cystine knot proteins.	BL01185D 23.45 8.043e-19 580-633
13	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 5.310e-11 391-419
13	PR00802	SERUM ALBUMIN FAMILY SIGNATURE	PR00802C 12.28 6.885e-10 470-488
16	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651 15.00 1.000e-10 43-56
17	BL00523	Sulfatases proteins.	BL00523E 19.27 8.125e-14 277-307 BL00523C 12.64 4.000e-13 145-156 BL00523A 13.36 7.300e-13 53-70 BL00523B 8.64 6.114e-11 99-111 BL00523D 9.89 2.174e-09 235-247
18	PF00023	Ank repeat proteins.	PF00023A 16.03 1.000e-11 181-197
18	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 2.273e-11 249-304 PF00791C 20.98 9.165e-10 263-302 PF00791B 28.49 9.650e-10 181-236 PF00791B 28.49 3.312e-09 148-203
19	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651 15.00 2.286e-10 46-59
21	PR00253	GAMMA-AMINOBUTYRIC ACID (GABA) RECEPTOR SIGNATURE	PR00253A 9.15 5.714e-24 246-267 PR00253C 13.85 9.500e-23 306-328 PR00253B 13.47 3.143e-22 272-294 PR00253D 16.68 7.000e-22 451-472
21	PR00252	NEUROTRANSMITTER-GATED ION CHANNEL FAMILY SIGNATURE	PR00252C 17.49 3.739e-15 161-176 PR00252A 14.28 4.115e-13 83-100 PR00252D 12.29 7.750e-11 237-250 PR00252B 15.17 3.250e-10 115-127
21	BL00236	Neurotransmitter-gated ion-channels proteins.	BL00236D 25.66 1.857e-32 230-272 BL00236C 25.16 2.227e-31 146-185 BL00236A 21.96 4.682e-24 63-101 BL00236B 14.67 9.625e-10 116-126

TABLE 4

SEQ ID	Model	Description	E-value	Score	Repeats	Position
12	EGF	EGF-like domain	2.3e-271	914.9	37	24-57:63-98:106-138:144-175:182-215:222-254:261-292:299-332:339-370:376-409:416-449:456-487:494-525:532-563:570-600:607-638:645-675:682-713:720-750:757-788:795-826:833-867:874-905:912-943:950-981:988-1019:1026-

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						1057:1064-1095:1111-1143:1150-1181:1188-1219:1226-1265:1272-1305:1312-1346:1353-1384:1392-1426:1529-1562
12	ank	Ankyrin repeat	5e-47	169.7	6	1881-1926:1928-1960:1961-1994:1995-2027:2028-2060:2061-2093
12	notch	Notch (DSL) domain	2.4e-39	144.2	3	1443-1481:1486-1523:1524-1563
12	disintegrin	Disintegrin	2.4	-25.5	1	811-863
12	Keratin B2	Keratin, high sulfur B2 protein	7.5	-82.0	1	103-258
12	Metallothio_PEC	Plant PEC family metallothionein	8.3	-40.6	1	1244-1307
13	vwc	von Willebrand factor type C domain	4.7e-06	33.6	3	309-378:418-481:486-548
13	Cys_knot	Cystine-knot domain	3.4e-05	26.7	1	541-653
13	vwd	von Willebrand factor type D domain	0.0009	18.2	1	8-133
13	TIL	Trypsin Inhibitor like cysteine rich domain	1.1	-4.2	1	254-309
16	BTB	BTB/POZ domain	1.9e-30	114.6	1	20-124
17	Sulfatase	Sulfatase	2.9e-94	326.6	1	53-479
18	ank	Ankyrin repeat	1.6e-48	174.6	5	143-175:176-208:209-241:243-276:277-309
18	fn3	Fibronectin type III domain	0.29	9.7	1	9-97
19	Kelch	Kelch motif	1.7e-36	134.7	5	348-399:401-449:451-497:499-544:546-600
19	BTB	BTB/POZ domain	8.4e-35	129.0	1	23-128
21	Neur_chan_LBD	Neurotransmitter-gated ion-channel lig	1.6e-81	284.3	1	37-243
21	Neur_chan_memb	Neurotransmitter-gated ion-channel tra	3.4e-67	236.7	1	250-468

Table 5

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
12	1a3d		1069	1219	1.9e-22	0.01	-0.09		PHOSPHOLIPASE A2; CHAIN: NULL;	CARBOXYLIC ESTER HYDROLASE PHOSPHOLIPASE, TRIMER, CALCIUM BINDING, ACTIVATOR SITE, 2 CARBOXYLIC ESTER HYDROLASE
12	1aut	L	1057	1158	7.6e-21	0.34	0.13		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
12	1aut	L	126	230	7.6e-25	0.37	0.76		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
12	1aut	L	1294	1398	5.7e-16	0.21	-0.03		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
12	1aut	L	738	846	5.7e-23	0.27	0.16		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
12	1aut	L	826	920	1.5e-22	0.31	0.84		ACTIVATED PROTEIN C; CHAIN: C, L; D-PHE-PRO-MAI; CHAIN: P;	COMPLEX (BLOOD COAGULATION/INHIBITOR) AUTOPROTHROMBIN IIA; HYDROLASE, SERINE PROTEINASE), PLASMA CALCIUM BINDING, 2 GLYCOPROTEIN, COMPLEX (BLOOD COAGULATION/INHIBITOR)
12	1awc	B	1929	2081	1.9e-41	0.32	1.00		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
12	1bd8		1874	2019	1.9e-26	0.46	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
12	1bd8		1929	2084	1.9e-40	0.16	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
12	1bi7	B	1995	2114	1.1e-28	0.34	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
12	1blx	B	1874	2054	7.6e-27	0.13	0.99		CYCLIN-DEPENDE	COMPLEX (INHIBITOR

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									NT KINASE 6; CHAIN: A; P19INK4D ; CHAIN: B;	PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
12	1blx	B	1931	2087	1.9e-39	-0.01	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D ; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
12	1blx	B	1962	2117	9.5e-36	0.44	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D ; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
12	1cej	A	1312	1384	9.5e-15	0.39	0.15		MEROZOITE SURFACE PROTEIN 1; CHAIN: A;	SURFACE PROTEIN MEROZOITE SURFACE ANTIGEN 1, MAJOR BLOOD-STAGE EGF-LIKE DOMAIN, EXTRACELLULAR, MODULAR PROTEIN, SURFACE 2 ANTIGEN, MALARIA VACCINE COMPONENT, SURFACE PROTEIN
12	1cej	A	791	908	3.8e-18	-0.05	0.06		MEROZOITE SURFACE PROTEIN 1; CHAIN: A;	SURFACE PROTEIN MEROZOITE SURFACE ANTIGEN 1, MAJOR BLOOD-STAGE EGF-LIKE DOMAIN, EXTRACELLULAR, MODULAR PROTEIN, SURFACE 2 ANTIGEN, MALARIA VACCINE COMPONENT, SURFACE PROTEIN
12	1cej	A	946	1028	1.9e-19	0.02	-0.18		MEROZOITE SURFACE	SURFACE PROTEIN MEROZOITE SURFACE ANTIGEN 1,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									PROTEIN 1; CHAIN: A;	MAJOR BLOOD-STAGE EGF-LIKE DOMAIN, EXTRACELLULAR, MODULAR PROTEIN, SURFACE 2 ANTIGEN, MALARIA VACCINE COMPONENT, SURFACE PROTEIN
12	1d9s	A	1988	2114	5.7e-34	0.07	0.95		CYCLIN-DEPENDENT KINASE 4 INHIBITOR B; CHAIN: A;	SIGNALING PROTEIN HELIX-TURN-HELIX, ANKYRIN REPEAT
12	1dan	L	1024	1163	7.6e-23	0.14	-0.05		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	1075	1201	7.6e-26	0.35	-0.02		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	1123	1237	9.5e-26	0.32	0.24		BLOOD COAGULATION FACTOR	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	1161	1285	7.6e-29	0.06	-0.12		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	125	227	1.5e-26	-0.15	0.43		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	150	274	3.8e-25	0.09	-0.01		BLOOD COAGULATION FACTOR VIIA;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	312	423	9.5e-26	0.22	-0.06		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	31	149	1.1e-22	0.20	0.23		BLOOD COAGULATION FACTOR VIIA; CHAIN: L, H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME, INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dan	L	505	620	3.8e-31	-0.32	0.00		BLOOD COAGULATION FACTOR VIIA; CHAIN: L,	BLOOD COAGULATION, SERINE PROTEASE, COMPLEX, CO-FACTOR, 2 RECEPTOR ENZYME,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									H; SOLUBLE TISSUE FACTOR; CHAIN: T, U; D-PHE-PHE-ARG-CHLORO METHYL KETONE (DFFRCMK) WITH CHAIN: C;	INHIBITOR, GLA, EGF, 3 COMPLEX (SERINE PROTEASE/COFACTOR/LIGAND)
12	1dcq	A	1922	2088	1.9e-29	-0.26	0.45		PYK2-ASSOCIATED PROTEIN BETA; CHAIN: A;	METAL BINDING PROTEIN ZINC-BINDING MODULE, ANKYRIN REPEATS, METAL BINDING PROTEIN
12	1dva	L	1019	1103	1.9e-20	0.25	0.31		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
12	1dva	L	1305	1398	5.7e-18	0.48	0.09		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									E-76; CHAIN: X, Y;	
12	1dva	L	1391	1516	1.9e-08	-0.00	-0.11		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
12	1dva	L	449	545	7.6e-26	-0.16	0.19		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
12	1dva	L	675	762	1.7e-26	0.31	0.53		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L,	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	
12	1dva	L	905	1001	7.6e-28	-0.00	0.31		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
12	1dva	L	943	1031	3.8e-27	0.28	0.12		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX
12	1dva	L	981	1069	1.1e-23	0.05	0.34		DES-GLA FACTOR VIIA (HEAVY CHAIN); CHAIN: H, I; DES-GLA	HYDROLASE/HYDROLASE INHIBITOR PROTEIN-PEPTIDE COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									FACTOR VIIA (LIGHT CHAIN); CHAIN: L, M; (DPN)-PHE-ARG; CHAIN: C, D; PEPTIDE E-76; CHAIN: X, Y;	
12	1dx5	I	1064	1181	1.9e-21	0.36	-0.08		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	1309	1427	1.3e-17	0.63	-0.12		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx	I	221	332	3.8e	0.19	0.03		THROMBI	SERINE PROTEINASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
	5				-29				N LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	411	532	5.7e-28	-0.17	0.40		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	492	600	5.7e-28	0.18	0.43		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBI	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									N INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	
12	1dx5	I	58	177	1.5e-25	0.17	-0.13		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	641	750	3.8e-26	0.07	0.49		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	717	826	3.8e-26	0.19	0.13		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN;

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	870	981	1.1e-27	0.06	0.01		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1dx5	I	945	1057	1.1e-29	0.20	0.21		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E,	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									F, G, H;	
12	1dx5	I	987	1092	1.9e-24	0.14	-0.06		THROMBIN LIGHT CHAIN; CHAIN: A, B, C, D; THROMBIN HEAVY CHAIN; CHAIN: M, N, O, P; THROMBOMODULIN; CHAIN: I, J, K, L; THROMBIN INHIBITOR L-GLU-L-GLY-L-ARM; CHAIN: E, F, G, H;	SERINE PROTEINASE COAGULATION FACTOR II; COAGULATION FACTOR II; FETOMODULIN, TM, CD141 ANTIGEN; EGR-CMK SERINE PROTEINASE, EGF-LIKE DOMAINS, ANTICOAGULANT COMPLEX, 2 ANTIFIBRINOLYTIC COMPLEX
12	1hre		978	1039	9.5e-19	0.17	0.01		GROWTH FACTOR HEREGULIN-ALPHA (EPIDERMAL GROWTH FACTOR-LIKE DOMAIN) 1HRE 3 (NMR, MINIMIZED AVERAGE STRUCTURE) 1HRE 4	
12	1ihb	A	1884	2052	7.6e-27	0.48	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
12	1ihb	A	1931	2084	9.5e-37	0.33	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
12	1klo		1276	1435	5.7e-24	0.17	-0.01		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
12	1klo		1316	1524	5.7e-19	0.04	-0.14		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		148	312	1.7e-31	0.30	-0.13		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		226	429	3.8e-30	0.20	0.01		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		306	467	1.9e-29	0.09	-0.06		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		31	195	3.8e-26	0.24	0.25		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		498	694	3.8e-32	0.22	0.35		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		686	846	1.7e-31	0.41	0.49		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		799	961	7.6e-32	0.13	0.05		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1klo		916	1085	1.1e-34	0.28	0.74		LAMININ; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN
12	1myo		1884	2013	1.5e-23	0.40	1.00		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
12	1myo		1930	2037	3.8e-25	0.14	0.46		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
12	1myo		1998	2112	1.9e-32	-0.08	0.74		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
12	1nfi	E	1881	2098	9.5e-50	0.42	1.00		NF-KAPPA-B P65; CHAIN: A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REG/ANK REPEAT) COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
12	1pfx	L	1066	1210	7.6e-39	0.17	0.00		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	1150	1297	3.8e-41	0.09	-0.14		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	1190	1337	1.9e-37	0.20	-0.05		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	1236	1375	1.5e-37	0.11	-0.03		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	1274	1417	9.5e-34	0.19	0.00		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										GLYCOPROTEIN
12	1pfx	L	146	283	5.7e-41	0.03	-0.09		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	189	323	1.9e-38	0.28	-0.12		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	222	361	5.7e-41	0.25	-0.06		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	378	516	5.7e-41	-0.11	0.03		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	458	592	3.8e-43	-0.35	0.09		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: I;	HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	65	207	1.9e-38	0.20	-0.13		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	836	972	5.7e-41	0.04	0.01		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	877	1010	1.9e-42	0.27	0.05		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	915	1048	5.7e-41	0.05	-0.06		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	952	1086	3.8e-40	0.17	0.46		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1qfk	L	1025	1161	5.7e-24	0.12	-0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	1117	1199	3.8e-26	0.28	0.25		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	1149	1246	3.8e-28	0.03	0.15		COAGULATION FACTOR VIIA (LIGHT	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	1187	1288	1.9e-29	0.11	0.60		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	1312	1404	1.9e-21	0.24	0.05		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	181	273	1.3e-27	0.56	0.21		COAGULATION FACTOR VIIA (LIGHT CHAIN);	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	298	387	9.5e-26	-0.13	0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	339	467	1.1e-22	0.05	0.00		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	415	508	9.5e-29	0.05	0.27		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	493	592	3.8e-29	-0.05	0.78		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	531	618	9.5e-30	0.46	0.37		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	63	155	1.7e-23	0.39	0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGUL	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									ATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	756	846	1.7e-29	0.35	0.78		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	912	1000	3.8e-31	-0.02	0.47		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	949	1037	3.8e-31	0.20	0.13		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	988	1075	3.8e-30	0.05	0.72		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1sw6	A	1874	2070	3.8e-26	0.05	0.72		REGULATORY PROTEIN SWI6; CHAIN: A, B;	TRANSCRIPTION REGULATION TRANSCRIPTION REGULATION, ANKYRIN REPEATS, CELL-CYCLE
12	1urk		1312	1458	3.8e-15	0.13	-0.14		PLASMINOGEN ACTIVATION PLASMINOGEN ACTIVATOR (UROKINASE-TYPE) (AMINO TERMINAL FRAGMENT) (NMR, 15 STRUCTURES)	
12	1urk		757	920	5.7e-17	0.07	-0.17		PLASMINOGEN ACTIVATION	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									PLASMINOGEN ACTIVATOR (UROKINASE-TYPE) (AMINO TERMINAL FRAGMENT) (NMR, 15 STRUCTURES)	
12	1vap	A	1059	1188	1.9e-30	0.08	-0.09		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	144	261	1.7e-29	0.16	-0.12		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	306	456	7.6e-28	0.01	-0.14		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	31	144	5.7e-23	0.04	-0.13		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	687	795	1.1e-28	0.00	-0.12		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vpi		106	222	9.5e-30	0.02	-0.18		PHOSPHOLIPASE A2 INHIBITOR; CHAIN: NULL	NEUROTOXIN PHOSPHOLIPASE A2 INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1vpi		1238	1353	9.5e-30	0.22	-0.17		PHOSPHOLIPASE A2 INHIBITOR; CHAIN: NULL	NEUROTOXIN PHOSPHOLIPASE A2 INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1vpi		988	1102	3.8e-31	0.13	-0.17		PHOSPHOLIPASE	NEUROTOXIN PHOSPHOLIPASE A2

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									A2 INHIBITOR; CHAIN: NULL	INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1whe		838	908	1.7e-16	0.20	-0.05		COAGULATION FACTOR X; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN, HYDROLASE, SERINE PROTEASE, PLASMA, BLOOD 2 COAGULATION FACTOR
12	1whe		995	1061	1.9e-19	0.36	-0.08		COAGULATION FACTOR X; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN, HYDROLASE, SERINE PROTEASE, PLASMA, BLOOD 2 COAGULATION FACTOR
12	1xka	L	1149	1248	9.5e-25	0.29	0.36		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	1312	1404	3.8e-18	0.29	0.13		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	144	233	3.8e-25	0.38	0.90		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	756	844	1.3e-26	0.33	0.48		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										GROWTH FACTOR LIKE DOMAIN
12	1xka	L	833	925	1.9e-24	0.24	0.53		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	949	1039	3.8e-28	0.15	0.99		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1ycs	B	1931	2093	1.9e-27	-0.17	0.64		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
12	1ycs	B	1998	2114	1.7e-28	0.10	0.87		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
12	2not	A	1117	1219	5.7e-31	0.05	-0.11		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										VENOM
12	2not	A	529	638	5.7e-30	0.16	-0.11		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	61	176	1.7e-25	-0.40	0.06		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	677	789	3.8e-32	-0.15	0.01		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	983	1095	5.7e-30	0.26	-0.09		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	9wga	A	1045	1257	1.9e-40	0.13	-0.12		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	1186	1375	1.1e-38	0.08	-0.06		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	1219	1419	1.9e-37	0.40	-0.11		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2)	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	189	323	1.9e-38	0.28	-0.12		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	222	361	5.7e-41	0.25	-0.06		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	378	516	5.7e-41	-0.11	0.03		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	Psi Blast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: I;	INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	458	592	3.8e-43	-0.35	0.09		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	65	207	1.9e-38	0.20	-0.13		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	836	972	5.7e-41	0.04	0.01		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	877	1010	1.9e-42	0.27	0.05		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	915	1048	5.7e-41	0.05	-0.06		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG;	COMPLEX (BLOOD COAGULATION/ INHIBITOR) CHRISTMAS FACTOR; COMPLEX,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: I;	INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1pfx	L	952	1086	3.8e-40	0.17	0.46		FACTOR IXA; CHAIN: C, L,; D-PHE-PRO-ARG; CHAIN: I;	COMPLEX (BLOOD COAGULATION/INHIBITOR) CHRISTMAS FACTOR; COMPLEX, INHIBITOR, HEMOPHILIA/EGF, BLOOD COAGULATION, 2 PLASMA, SERINE PROTEASE, CALCIUM-BINDING, HYDROLASE, 3 GLYCOPROTEIN
12	1qfk	L	1025	1161	5.7e-24	0.12	-0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAV	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									Y CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	1117	1199	3.8e-26	0.28	0.25		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	1149	1246	3.8e-28	0.03	0.15		COAGULATION FACTOR VIIA (LIGHT CHAIN)	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									; CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	1187	1288	1.9e-29	0.11	0.60		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: C;	
12	1qfk	L	1312	1404	1.9e-21	0.24	0.05		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	181	273	1.3e-27	0.56	0.21		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAV	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									Y CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	298	387	9.5e-26	-0.13	0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	339	467	1.1e-22	0.05	0.00		COAGULATION FACTOR VIIA (LIGHT CHAIN)	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									; CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	415	508	9.5e-29	0.05	0.27		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: C;	
12	1qfk	L	493	592	3.8e-29	-0.05	0.78		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	531	618	9.5e-30	0.46	0.37		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAV	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									Y CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	63	155	1.7e-23	0.39	0.11		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	756	846	1.7e-29	0.35	0.78		COAGULATION FACTOR VIIA (LIGHT CHAIN)	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									; CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1qfk	L	912	1000	3.8e-31	-0.02	0.47		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: C;	
12	1qfk	L	949	1037	3.8e-31	0.20	0.13		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
12	1qfk	L	988	1075	3.8e-30	0.05	0.72		COAGULATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAV	SERINE PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									Y CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	
12	1sw6	A	1874	2070	3.8e-26	0.05	0.72		REGULATORY PROTEIN SWI6; CHAIN: A, B;	TRANSCRIPTION REGULATION TRANSCRIPTION REGULATION, ANKYRIN REPEATS, CELL-CYCLE
12	1urk		1312	1458	3.8e-15	0.13	-0.14		PLASMINOGEN ACTIVATION PLASMINOGEN ACTIVATOR (UROKINASE-TYPE) (AMINO TERMINAL FRAGMENT) (NMR, 15 STRUCTURES)	
12	1urk		757	920	5.7e-17	0.07	-0.17		PLASMINOGEN ACTIV	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									ATION PLASMINOGEN ACTIVATOR (UROKINASE-TYPE) (AMINO TERMINAL FRAGMENT) (NMR, 15 STRUCTURES)	
12	1vap	A	1059	1188	1.9e-30	0.08	-0.09		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	144	261	1.7e-29	0.16	-0.12		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	306	456	7.6e-28	0.01	-0.14		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	31	144	5.7e-23	0.04	-0.13		PHOSPHOLIPASE A2; CHAIN: A, B;	LIPID DEGRADATION PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vap	A	687	795	1.1e-	0.00	-0.12		PHOSPHOLIP	LIPID DEGRADATION

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
					28				ASE A2; CHAIN: A, B;	PHOSPHOLIPASE A2, LIPID DEGRADATION, HYDROLASE
12	1vpi		106	222	9.5e-30	0.02	-0.18		PHOSPHOLIPASE A2 INHIBITOR; CHAIN: NULL	NEUROTOXIN PHOSPHOLIPASE A2 INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1vpi		1238	1353	9.5e-30	0.22	-0.17		PHOSPHOLIPASE A2 INHIBITOR; CHAIN: NULL	NEUROTOXIN PHOSPHOLIPASE A2 INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1vpi		988	1102	3.8e-31	0.13	-0.17		PHOSPHOLIPASE A2 INHIBITOR; CHAIN: NULL	NEUROTOXIN PHOSPHOLIPASE A2 INHIBITOR, X-RAY STRUCTURE, RECOGNITION, 2 MOLECULAR EVOLUTION, NEUROTOXIN
12	1whe		838	908	1.7e-16	0.20	-0.05		COAGULATION FACTOR X; CHAIN: NULL;	GLYCOPROTEIN GLYCOPROTEIN, HYDROLASE, SERINE PROTEASE, PLASMA, BLOOD 2 COAGULATION FACTOR
12	1whe		995	1061	1.9e-19	0.36	-0.08		COAGULATION FACTOR X;	GLYCOPROTEIN GLYCOPROTEIN, HYDROLASE, SERINE PROTEASE,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: NULL;	PLASMA, BLOOD 2 COAGULATION FACTOR
12	1xka	L	1149	1248	9.5e-25	0.29	0.36		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	1312	1404	3.8e-18	0.29	0.13		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	144	233	3.8e-25	0.38	0.90		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										DOMAIN
12	1xka	L	756	844	1.3e-26	0.33	0.48		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	833	925	1.9e-24	0.24	0.53		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1xka	L	949	1039	3.8e-28	0.15	0.99		BLOOD COAGULATION FACTOR XA; CHAIN: L, C;	BLOOD COAGULATION FACTOR STUART FACTOR; BLOOD COAGULATION FACTOR, SERINE PROTEINASE, EPIDERMAL 2 GROWTH FACTOR LIKE DOMAIN
12	1ycs	B	1931	2093	1.9e-	-0.17	0.64		P53; CHAIN:	COMPLEX (ANTI-

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
					27				A; 53BP2; CHAIN: B;	ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
12	1ycs	B	1998	2114	1.7e-28	0.10	0.87		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3 POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
12	2n	A	111	12	5.7	0.05	-		PHOSP	HYDROLASE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
	ot		7	19	e-31		0.11		HOLIPASE A2; CHAIN: A, B;	HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	529	638	5.7e-30	0.16	-0.11		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	61	176	1.7e-25	-0.40	0.06		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	677	789	3.8e-32	-0.15	0.01		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	2not	A	983	1095	5.7e-30	0.26	-0.09		PHOSPHOLIPASE A2; CHAIN: A, B;	HYDROLASE HYDROLASE, LIPID DEGRADATION, CALCIUM, PRESYNAPTIC 2 NEUROTOXIN, VENOM
12	9wga	A	1045	1257	1.9e-40	0.13	-0.12		LECTIN (AGGLUTININ) WHEAT	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	1186	1375	1.1e-38	0.08	-0.06		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	1219	1419	1.9e-37	0.40	-0.11		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	1346	1552	3.8e-19	0.05	0.01		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									3	
12	9wga	A	142	323	7.6e-45	0.39	0.18		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	215	402	1.9e-44	0.54	-0.11		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	24	208	1.5e-36	0.32	-0.11		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	374	554	5.7e-44	0.07	0.03		LECTIN (AGGLUTININ) WHEA	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									T GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	448	629	3.8e-42	0.01	-0.13		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	563	742	7.6e-44	0.14	-0.12		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA 3	
12	9wga	A	56	245	1.7e-43	0.13	-0.13		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2)	

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									9WGA3	
12	9wga	A	680	858	1.9e-44	0.25	0.17		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA3	
12	9wga	A	826	1011	1.7e-45	0.17	-0.03		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA3	
12	9wga	A	905	1088	3.8e-48	0.24	-0.06		LECTIN (AGGLUTININ) WHEAT GERM AGGLUTININ (ISOLECTIN 2) 9WGA3	
13	1bmp		568	632	0.00076	-0.47	0.35		BONE MORPHOGENE	TRANSFORMING GROWTH FACTOR

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									TIC PROTEIN-7; CHAIN: NULL;	OSTEOGENIC PROTEIN-1, HOP-1, BMP-7; MORPHOGEN, TRANSFORMING GROWTH FACTOR, CYTOKINE, BONE, 2 CARTILAGE, GLYCOPROTEIN
13	1ccv	A	254	309	9.5e-12	-0.03	0.11		APIS MELLIFERA CHYMOTRYPSIN INHIBITOR; CHAIN: A;	PROTEIN INHIBITOR AMCI PROTEIN INHIBITOR, HEMOLYMPH, APIS MELLIFERA, CANONICAL 2 INHIBITOR
13	1cou	A	252	322	3.8e-12	0.20	0.06		NEMATODE ANTICOAGULANT PROTEIN C2; CHAIN: A;	BLOOD CLOTTING NAPC2; ANTICOAGULANT, PROTEASE INHIBITOR, BLOOD CLOTTING
13	1eai	C	251	309	1.9e-14	-0.12	0.06		ELASTASE; CHAIN: A, B; CHYMOTRYPSIN/ELASTASE ISOINHIBITOR 1; CHAIN: C, D;	SERINE PROTEINASE SERINE PROTEINASE, ELASTASE, ASCARIS SUMM, PROTEIN INHIBITOR
13	1q	L	368	46	1.9	-	0.00		COAGU	SERINE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
	fk			4	e-05	0.23			LATION FACTOR VIIA (LIGHT CHAIN); CHAIN: L; COAGULATION FACTOR VIIA (HEAVY CHAIN); CHAIN: H; TRIPEPTIDYL INHIBITOR; CHAIN: C;	PROTEASE FVIIA; FVIIA; BLOOD COAGULATION, SERINE PROTEASE
16	1buo	A	2	124	1.5e-22			57.90	PROMYELOCYTIC LEUKEMIA ZINC FINGER PROTEIN PLZF; CHAIN: A;	GENE REGULATION POZ DOMAIN; PROTEIN-PROTEIN INTERACTION DOMAIN, TRANSCRIPTIONAL 2 REPRESSOR, ZINC-FINGER PROTEIN, X-RAY CRYSTALLOGRAPHY, 3 PROTEIN STRUCTURE, PROMYELOCYT

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										IC LEUKEMIA, GENE REGULATION
16	1buo	A	5	96	1.5e-22	0.22	1.00		PROMYELOCYTIC LEUKEMIA ZINC FINGER PROTEIN PLZF; CHAIN: A;	GENE REGULATION POZ DOMAIN; PROTEIN-PROTEIN INTERACTION DOMAIN, TRANSCRIPTIONAL 2 REPRESSOR, ZINC-FINGER PROTEIN, X-RAY CRYSTALLOGRAPHY, 3 PROTEIN STRUCTURE, PROMYELOCYTIC LEUKEMIA, GENE REGULATION
16	1cex		188	283	0.0095	0.52	0.05		CUTINASE; CHAIN: NULL;	SERINE ESTERASE HYDROLASE, SERINE ESTERASE, GLYCOPROTEIN
17	1auk		51	541	1.9e-84			224.76	ARYLSULFATASE A; CHAIN: NULL;	HYDROLASE CEREBROSIDE-3-SULFATE-SULFATASE; CEREBROSIDE-3-SULFATE HYDROLYSIS, LYSOSOMAL ENZYME, 2 HYDROLASE
17	1auk		53	522	1.9e-84	0.58	1.00		ARYLSULFATASE A; CHAIN:	HYDROLASE CEREBROSIDE-3-SULFATE-SULFATASE;

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									NULL;	CEREBROSIDE-3-SULFATE HYDROLYSIS, LYSOSOMAL ENZYME, 2 HYDROLASE
18	1a5e		114	227	7.6e-22	0.35	-0.02		TUMOR SUPPRESSOR P16INK4A; CHAIN: NULL;	ANTI-ONCOGENE CELL CYCLE, ANTI-ONCOGENE, REPEAT, ANK REPEAT
18	1a5e		130	261	3.8e-24	0.44	0.76		TUMOR SUPPRESSOR P16INK4A; CHAIN: NULL;	ANTI-ONCOGENE CELL CYCLE, ANTI-ONCOGENE, REPEAT, ANK REPEAT
18	1a5e		163	294	1.9e-25	0.62	0.77		TUMOR SUPPRESSOR P16INK4A; CHAIN: NULL;	ANTI-ONCOGENE CELL CYCLE, ANTI-ONCOGENE, REPEAT, ANK REPEAT
18	1a5e		196	308	1.5e-27	0.61	1.00		TUMOR SUPPRESSOR P16INK4A; CHAIN: NULL;	ANTI-ONCOGENE CELL CYCLE, ANTI-ONCOGENE, REPEAT, ANK REPEAT
18	1awc	B	114	197	1.1e-17	0.62	1.00		GA BINDING PROTEIN ALPHA	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA;

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
18	1awc	B	146	297	1.9e-40	0.86	1.00		GA BINDING PROTEIN ALPHA ; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
18	1awc	B	179	330	1.9e-40			59.47	GA BINDING PROTEIN ALPHA ; CHAIN: A; GA BINDING	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/D

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	NA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
18	1awc	B	181	318	1.9e-36	0.43	1.00		GA BINDING PROTEIN ALPHA; CHAIN: A; GA BINDING PROTEIN BETA 1; CHAIN: B; DNA; CHAIN: D, E;	COMPLEX (TRANSCRIPTION REGULATION/DNA) GABPALPHA; GABPBETA1; COMPLEX (TRANSCRIPTION REGULATION/DNA), DNA-BINDING, 2 NUCLEAR PROTEIN, ETS DOMAIN, ANKYRIN REPEATS, TRANSCRIPTION 3 FACTOR
18	1bd8		109	267	1.9e-39			53.03	P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
18	1bd8		114	199	1.3e-16	0.43	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
18	1bd8		144	299	1.9e-39	0.78	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
18	1bd8		210	318	9.5e-25	0.61	1.00		P19INK4D CDK4/6 INHIBITOR; CHAIN: NULL;	TUMOR SUPPRESSOR TUMOR SUPPRESSOR, CDK4/6 INHIBITOR, ANKYRIN MOTIF
18	1bi7	B	114	196	5.7e-15	0.55	0.99		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
18	1bi7	B	137	230	1.9e-21	0.83	0.99		CYCLIN-DEPENDENT KINASE 6;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
18	1bi7	B	144	299	7.6e-26	0.38	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A, MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
18	1bi7	B	210	318	1.9e-20	0.41	1.00		CYCLIN-DEPENDENT KINAS	COMPLEX (KINASE/ANTI-ONCOGENE) CDK6; P16INK4A,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									E 6; CHAIN: A; MULTIPLE TUMOR SUPPRESSOR; CHAIN: B;	MTS1; CYCLIN DEPENDENT KINASE, CYCLIN DEPENDENT KINASE INHIBITORY 2 PROTEIN, CDK, INK4, CELL CYCLE, MULTIPLE TUMOR SUPPRESSOR, 3 MTS1, COMPLEX (KINASE/ANTI-ONCOGENE) HEADER
18	1blx	B	109	269	7.6e-39			52.22	CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
18	1blx	B	146	303	7.6e-39	0.83	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										(INHIBITOR PROTEIN/KINASE)
18	1blx	B	179	318	5.7e-32	0.49	1.00		CYCLIN-DEPENDENT KINASE 6; CHAIN: A; P19INK4D; CHAIN: B;	COMPLEX (INHIBITOR PROTEIN/KINASE) INHIBITOR PROTEIN, CYCLIN-DEPENDENT KINASE, CELL CYCLE 2 CONTROL, ALPHA/BETA, COMPLEX (INHIBITOR PROTEIN/KINASE)
18	1bpv		8	111	1.1e-09	0.45	0.07		TITIN; CHAIN: NULL;	CONNECTIN A71, CONNECTIN; TITIN, CONNECTIN, FIBRONECTIN TYPE III
18	1bu9	A	139	309	3.8e-37			55.75	CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN- 2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
18	1bu9	A	146	309	3.8e-37	0.65	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBI	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									TOR; CHAIN: A;	SUPPRESSOR, CYCLIN- 2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
18	1bu9	A	176	323	1.9e-30	0.65	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A;	HORMONE/GROWTH FACTOR P18-INK4C; CELL CYCLE INHIBITOR, P18INK4C, TUMOR, SUPPRESSOR, CYCLIN- 2 DEPENDENT KINASE, HORMONE/GROWTH FACTOR
18	1c8p	A	11	104	1.5e-05	0.15	0.00		CYTOKINE RECEPTOR COMMON BETA CHAIN; CHAIN: A;	MEMBRANE PROTEIN BETA SANDWICH, CYTOKINE RECEPTOR, FN3 DOMAIN
18	1cd9	B	11	105	3.8e-05	0.38	0.37		GRANULOCYTE COLONY-STIMULATING FACTOR; CHAIN: A, C; G-CSF RECEPTOR; CHAIN:	CYTOKINE G-CSF; G-CSF-R; CLASS1 CYTOKINE, HEMATOPOIETIC RECEPTOR, SIGNAL TRANSDUCTION

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									B, D;	
18	1d9s	A	114	233	7.6e-26	1.07	1.00		CYCLIN-DEPENDENT KINASE 4 INHIBITOR B; CHAIN: A;	SIGNALING PROTEIN HELIX-TURN-HELIX, ANKYRIN REPEAT
18	1d9s	A	169	303	3.8e-38	0.59	1.00		CYCLIN-DEPENDENT KINASE 4 INHIBITOR B; CHAIN: A;	SIGNALING PROTEIN HELIX-TURN-HELIX, ANKYRIN REPEAT
18	1dcq	A	113	233	5.7e-22	0.97	1.00		PYK2-ASSOCIATED PROTEIN BETA; CHAIN: A;	METAL BINDING PROTEIN ZINC-BINDING MODULE, ANKYRIN REPEATS, METAL BINDING PROTEIN
18	1dcq	A	136	302	3.8e-28	0.14	1.00		PYK2-ASSOCIATED PROTEIN BETA; CHAIN: A;	METAL BINDING PROTEIN ZINC-BINDING MODULE, ANKYRIN REPEATS, METAL BINDING PROTEIN
18	1ihb	A	145	301	1.9e-36			53.04	CYCLIN-DEPENDENT	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									KINASE 6 INHIBITOR; CHAIN: A, B;	INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
18	1ihb	A	146	301	1.9e-36	0.81	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
18	1ihb	A	181	318	5.7e-30	0.53	1.00		CYCLIN-DEPENDENT KINASE 6 INHIBITOR; CHAIN: A, B;	CELL CYCLE INHIBITOR P18-INK4C(INK6); CELL CYCLE INHIBITOR, P18-INK4C(INK6), ANKYRIN REPEAT, 2 CDK 4/6 INHIBITOR
18	1myo		114	227	1.1e-27	0.67	1.00		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
18	1myo		181	295	5.7e-32	0.48	1.00		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
18	1myo		212	318	9.5e-29	0.32	0.99		MYOTROPHIN; CHAIN: NULL	ANK-REPEAT MYOTROPHIN, ACETYLATION, NMR, ANK-REPEAT
18	1nfi	E	106	309	1.9e-49			58.68	NF-KAPPA-B P65; CHAIN:	COMPLEX (TRANSCRIPTION REG/ANK REPEAT)

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
18	1nfi	E	114	309	1.9e-49	0.45	1.00		NF-KAPPA-B P65; CHAIN: A, C; NF-KAPPA-B P50; CHAIN: B, D; I-KAPPA-B-ALPHA; CHAIN: E, F;	COMPLEX (TRANSCRIPTION REG/ANK REPEAT) COMPLEX (TRANSCRIPTION REGULATION/ANK REPEAT), ANKYRIN 2 REPEAT HELIX
18	1ycs	B	146	319	1.9e-36	0.34	1.00		P53; CHAIN: A; 53BP2; CHAIN: B;	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS) P53BP2; ANKYRIN REPEATS, SH3, P53, TUMOR SUPPRESSOR, MULTIGENE 2 FAMILY, NUCLEAR PROTEIN, PHOSPHORYLATION, DISEASE MUTATION, 3

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
										POLYMORPHISM, COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)
18	2fnb	A	8	102	1.9e-07	0.38	0.47		FIBRONECTIN; CHAIN: A;	PROTEIN BINDING ED-B, FIBRONECTIN, TYPEIII DOMAIN, ANGIOGENESIS, PROTEIN 2 BINDING
18	3hhr	B	3	104	1.1e-06	0.20	0.00		HORMONE/RECEPTOR HUMAN GROWTH HORMONE COMPLEXED WITH ITS RECEPTOR 3HHR 3 (EXTRACELLULAR DOMAIN) 3HHR 4	
19	1buo	A	7	124	1.7e-26	0.38	0.99		PROMYELOCYTIC LEUKEMIA ZINC FINGE	GENE REGULATION POZ DOMAIN; PROTEIN-PROTEIN INTERACTION DOMAIN,

SEQ ID NO:	PDB ID	CHAIN ID	START AA	END AA	PsiBlast	Verify score	PMF score	SEQFOLD score	Coumpound	PDB annotation
									R PROTEIN PLZF; CHAIN: A;	TRANSCRIPTIONAL 2 REPRESSOR, ZINC-FINGER PROTEIN, X-RAY CRYSTALLOGRAPHY, 3 PROTEIN STRUCTURE, PROMYELOCYTIC LEUKEMIA, GENE REGULATION
19	1gof		411	495	5.1e-11	0.07	0.34		OXIDOREDUCTASE(OXYGEN(A)) GALACTOSE OXIDASE (E.C.1.1.3.9) (PH 4.5) 1GOF 3	

TABLE 6

SEQ ID	Position	Maximum score	Average score
12	18	0.983	0.962
17	26	0.993	0.883
21	22	0.973	0.875

TABLE 7

SEQ ID	Chromsomal location
1	2
2	11p15
3	3p21.2-24.2
4	22
6	4
7	Xq21.3-q22
8	15
9	10cen-q26.11
10	15q11.2-q12

TABLE 8

SEQ ID NO: of Full-length Nucleotide Sequence	SEQ ID NO: of Full-length Pepttide Sequence	SEQ ID NO: in Priority Application USSN 09/815,925
1	12	1
2	13	2
3	14	3
4	15	4
5	16	5
6	17	6
7	18	7
8	19	8
9	20	9
10	21	10
11	22	11

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 – 11, a mature protein coding portion of SEQ ID NO: 1 – 11, an active domain coding protein of SEQ ID NO: 1 – 11, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

3. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

4. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

5. A vector comprising the polynucleotide of claim 1.

6. An expression vector comprising the polynucleotide of claim 1.

7. A host cell genetically engineered to comprise the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

9. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of a polypeptide encoded by any one of the polynucleotides of claim 1.

10. A composition comprising the polypeptide of claim 9 and a carrier.

11. An antibody directed against the polypeptide of claim 9.

12. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

14. The method of claim 13, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

15. A method for detecting the polypeptide of claim 9 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 9 is detected.

16. A method for identifying a compound that binds to the polypeptide of claim 9, comprising:

a) contacting the compound with the polypeptide of claim 9 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 9 is identified.

17. A method for identifying a compound that binds to the polypeptide of claim 9, comprising:

a) contacting the compound with the polypeptide of claim 9, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 9 is identified.

18. A method of producing the polypeptide of claim 9, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-11, a mature protein coding portion of SEQ ID NO: 1-11, an active domain coding portion of SEQ ID NO: 1-11, complementary sequences thereof, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

19. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides from the Sequence Listing, the mature protein portion thereof, or the active domain thereof.

20. The polypeptide of claim 21 wherein the polypeptide is provided on a polypeptide array.

21. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO: 1 – 11.

22. The collection of claim 21, wherein the collection is provided on a nucleic acid array.

23. The collection of claim 22, wherein the array detects full-matches to any one of the polynucleotides in the collection.

24. The collection of claim 22, wherein the array detects mismatches to any one of the polynucleotides in the collection.

25. The collection of claim 21, wherein the collection is provided in a computer-readable format.

26. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 9 or 19 and a pharmaceutically acceptable carrier.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 9 or 19 and a pharmaceutically acceptable carrier.

SEQUENCE LISTING

<110> Tang, Y. Tom
Zhou, Ping
Goodrich, Ryle
Asundi, Vinod
Yang, Yonghong
Zhang, Jie
Wehrman, Tom
Drmanac, Radoje T.

<120> Novel Nucleic Acids and Polypeptides

<130> 21272-029CIP2H-PCT

<140> Not Yet Assigned
<141> 2002-03-20

<150> 09/815,925

<151> 2001-03-22

<160> 11

<170> PatentIn version 3.0

<210> 1
<211> 7673
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(7671)

<400> 1
atg ccg ccg ctc ctg gcg ccc ctg ctc tgc ctg gcg ctg ctg ccc gcg 48
Met Pro Pro Leu Leu Ala Pro Leu Leu Cys Leu Ala Leu Leu Pro Ala
1 5 10 15

ctc gcc gca cga ggc ccg cga tgc tcc cag ccc ggt gag acc tgc ctg 96
Leu Ala Ala Arg Gly Pro Arg Cys Ser Gln Pro Gly Glu Thr Cys Leu

20 25 30

aat ggc ggg aag tgt gaa gcg gcc aat ggc acg gag gcc tgc gtc tgt 144
Asn Gly Gly Lys Cys Glu Ala Ala Asn Gly Thr Glu Ala Cys Val Cys
35 40 45

ggc ggg gcc ttc gtg ggc ccg cga tgc cag gac ccc aac ccg tgc ctc 192
Gly Gly Ala Phe Val Gly Pro Arg Cys Gln Asp Pro Asn Pro Cys Leu
50 55 60

agc acc ccc tgc aag aac gcc ggg aca tgc cac gtg gtg gac cgc aga 240
Ser Thr Pro Cys Lys Asn Ala Gly Thr Cys His Val Val Asp Arg Arg
65 70 75 80

ggc gtg gca gac tat gcc tgc agc tgt gcc ctg ggc ttc tct ggg ccc 288
Gly Val Ala Asp Tyr Ala Cys Ser Cys Ala Leu Gly Phe Ser Gly Pro
85 90 95

ctc tgc ctg aca ccc ctg gac aac gcc tgc ctc acc aac ccc tgc cgc 336
Leu Cys Leu Thr Pro Leu Asp Asn Ala Cys Leu Thr Asn Pro Cys Arg
100 105 110

aac ggg ggc acc tgc gac ctg ctc acg ctg acg gag tac aag tgc cgc 384
Asn Gly Gly Thr Cys Asp Leu Leu Thr Leu Thr Glu Tyr Lys Cys Arg
115 120 125

tgc ccg ccc ggc tgg tca ggg aaa tcg tgc cag cag gct gac ccg tgc 432
Cys Pro Pro Gly Trp Ser Gly Lys Ser Cys Gln Gln Ala Asp Pro Cys
130 135 140

gcc tcc aac ccc tgc gcc aac ggt ggc cag tgc ctg ccc ttc gag gcc 480
Ala Ser Asn Pro Cys Ala Asn Gly Gly Gln Cys Leu Pro Phe Glu Ala
145 150 155 160

tcc tac atc tgc cac tgc cca ccc agc ttc cat ggc ccc acc tgc cgg 528
Ser Tyr Ile Cys His Cys Pro Pro Ser Phe His Gly Pro Thr Cys Arg
165 170 175

cag gat gtc aac gag tgt ggc cag aag ccc agg ctt tgc cgc cac gga 576
Gln Asp Val Asn Glu Cys Gly Gln Lys Pro Arg Leu Cys Arg His Gly
180 185 190

ggc acc tgc cac aac gag gtc ggc tcc tac cgc tgc gtc tgc cgc gcc 624
Gly Thr Cys His Asn Glu Val Gly Ser Tyr Arg Cys Val Cys Arg Ala
195 200 205

acc cac act ggc ccc aac tgc gag cgg ccc tac gtg ccc tgc agc ccc 672
Thr His Thr Gly Pro Asn Cys Glu Arg Pro Tyr Val Pro Cys Ser Pro
210 215 220

tcg ccc tgc cag aac ggg ggc acc tgc cgc ccc acg ggc gac gtc acc 720
Ser Pro Cys Gln Asn Gly Gly Thr Cys Arg Pro Thr Gly Asp Val Thr
225 230 235 240

cac gag tgt gcc tgc ctg cca ggc ttc acc ggc cag aac tgt gag gaa 768
His Glu Cys Ala Cys Leu Pro Gly Phe Thr Gly Gln Asn Cys Glu Glu
245 250 255

aat atc gac gat tgt cca gga aac aac tgc aag aac ggg ggt gcc tgt 816
Asn Ile Asp Asp Cys Pro Gly Asn Asn Cys Lys Asn Gly Gly Ala Cys
260 265 270

gtg gac ggc gtg aac acc tac aac tgc ccg tgc ccg cca gag tgg aca 864
Val Asp Gly Val Asn Thr Tyr Asn Cys Pro Cys Pro Pro Glu Trp Thr
275 280 285

ggt cag tac tgt acc gag gat gtg gac gag tgc cag ctg atg cca aat 912

Gly Gln Tyr Cys Thr Glu Asp Val Asp Glu Cys Gln Leu Met Pro Asn
290 295 300

gcc tgc cag aac ggc ggg acc tgc cac aac acc cac ggt ggc tac aac 960
Ala Cys Gln Asn Gly Gly Thr Cys His Asn Thr His Gly Gly Tyr Asn
305 310 315 320

tgc gtg tgt gtc aac ggc tgg act ggt gag gac tgc agc gag aac att 1008
Cys Val Cys Val Asn Gly Trp Thr Gly Glu Asp Cys Ser Glu Asn Ile
325 330 335

gat gac tgt gcc agc gcc gcc tgc ttc cac ggc gcc acc tgc cat gac 1056
Asp Asp Cys Ala Ser Ala Ala Cys Phe His Gly Ala Thr Cys His Asp
340 345 350

cgt gtg gcc tcc ttt tac tgc gag tgt ccc cat ggc cgc aca ggt ctg 1104
Arg Val Ala Ser Phe Tyr Cys Glu Cys Pro His Gly Arg Thr Gly Leu
355 360 365

ctg tgc cac ctc aac gac gca tgc atc agc aac ccc tgt aac gag ggc 1152
Leu Cys His Leu Asn Asp Ala Cys Ile Ser Asn Pro Cys Asn Glu Gly
370 375 380

tcc aac tgc gac acc aac cct gtc aat ggc aag gcc atc tgc acc tgc 1200
Ser Asn Cys Asp Thr Asn Pro Val Asn Gly Lys Ala Ile Cys Thr Cys
385 390 395 400

ccc tcg ggg tac acg ggc ccg gcc tgc agc cag gac gtg gat gag tgc 1248
Pro Ser Gly Tyr Thr Gly Pro Ala Cys Ser Gln Asp Val Asp Glu Cys
405 410 415

tcg ctg ggt gcc aac ccc tgc gag cat gcg ggc aag tgc atc aac acg 1296
Ser Leu Gly Ala Asn Pro Cys Glu His Ala Gly Lys Cys Ile Asn Thr

420 425 430

ctg ggc tcc ttc gag tgc cag tgt ctg cag ggc tac acg ggc ccc cga 1344
Leu Gly Ser Phe Glu Cys Gln Cys Leu Gln Gly Tyr Thr Gly Pro Arg
435 440 445

tgc gag atc gac gtc aac gag tgc gtc tcg aac ccg tgc cag aac gac 1392
Cys Glu Ile Asp Val Asn Glu Cys Val Ser Asn Pro Cys Gln Asn Asp
450 455 460

gcc acc tgc ctg gac cag att ggg gag ttc cag tgc atg tgc atg ccc 1440
Ala Thr Cys Leu Asp Gln Ile Gly Glu Phe Gln Cys Met Cys Met Pro
465 470 475 480

ggc tac gag ggt gtg cac tgc gag gtc aac aca gac gag tgt gcc agc 1488
Gly Tyr Glu Gly Val His Cys Glu Val Asn Thr Asp Glu Cys Ala Ser
485 490 495

agc ccc tgc ctg cac aat ggc cgc tgc ctg gac aag atc aat gag ttc 1536
Ser Pro Cys Leu His Asn Gly Arg Cys Leu Asp Lys Ile Asn Glu Phe
500 505 510

cag tgc gag tgc ccc acg ggc ttc act ggg cat ctg tgc cag tac gat 1584
Gln Cys Glu Cys Pro Thr Gly Phe Thr Gly His Leu Cys Gln Tyr Asp
515 520 525

gtg gac gag tgt gcc agc acc ccc tgc aag aat ggt gcc aag tgc ctg 1632
Val Asp Glu Cys Ala Ser Thr Pro Cys Lys Asn Gly Ala Lys Cys Leu
530 535 540

gac gga ccc aac act tac acc tgt gtg tgc acg gaa ggg tac acg ggg 1680
Asp Gly Pro Asn Thr Tyr Thr Cys Val Cys Thr Glu Gly Tyr Thr Gly
545 550 555 560

acg cac tgc gag gtg gac atc gat gag tgc gac ccc gac ccc tgc cac 1728
Thr His Cys Glu Val Asp Ile Asp Glu Cys Asp Pro Asp Pro Cys His
565 570 575

tac ggc tcc tgc aag gac ggc gtc gcc acc ttc acc tgc ctc tgc cgc 1776
Tyr Gly Ser Cys Lys Asp Gly Val Ala Thr Phe Thr Cys Leu Cys Arg
580 585 590

cca ggc tac acg ggc cac cac tgc gag acc aac atc aac gag tgc tcc 1824
Pro Gly Tyr Thr Gly His His Cys Glu Thr Asn Ile Asn Glu Cys Ser
595 600 605

agc cag ccc tgc cgc cta cgg ggc acc tgc cag gac ccg gac aac gcc 1872
Ser Gln Pro Cys Arg Leu Arg Gly Thr Cys Gln Asp Pro Asp Asn Ala
610 615 620

tac ctc tgc ttc tgc ctg aag ggg acc aca gga ccc aac tgc gag atc 1920
Tyr Leu Cys Phe Cys Leu Lys Gly Thr Thr Gly Pro Asn Cys Glu Ile
625 630 635 640

aac ctg gat gac tgt gcc agc agc ccc tgc gac tcg ggc acc tgt ctg 1968
Asn Leu Asp Asp Cys Ala Ser Ser Pro Cys Asp Ser Gly Thr Cys Leu
645 650 655

gac aag atc gat ggc tac gag tgt gcc tgt gag ccg ggc tac aca ggg 2016
Asp Lys Ile Asp Gly Tyr Glu Cys Ala Cys Glu Pro Gly Tyr Thr Gly
660 665 670

agc atg tgt aac agc aac atc gat gag tgt gcg ggc aac ccc tgc cac 2064
Ser Met Cys Asn Ser Asn Ile Asp Glu Cys Ala Gly Asn Pro Cys His
675 680 685

aac ggg ggc acc tgc gag gac ggc atc aat ggc ttc acc tgc cgc tgc 2112

Asn Gly Gly Thr Cys Glu Asp Gly Ile Asn Gly Phe Thr Cys Arg Cys
690 695 700

ccc gag ggc tac cac gac ccc acc tgc ctg tct gag gtc aat gag tgc 2160
Pro Glu Gly Tyr His Asp Pro Thr Cys Leu Ser Glu Val Asn Glu Cys
705 710 715 720

aac agc aac ccc tgc gtc cac ggg gcc tgc cgg gac agc ctc aac ggg 2208
Asn Ser Asn Pro Cys Val His Gly Ala Cys Arg Asp Ser Leu Asn Gly
725 730 735

tac aag tgc gac tgt gac cct ggg tgg agt ggg acc aac tgt gac atc 2256
Tyr Lys Cys Asp Cys Asp Pro Gly Trp Ser Gly Thr Asn Cys Asp Ile
740 745 750

aac aac aac gag tgt gaa tcc aac cct tgt gtc aac ggc ggc acc tgc 2304
Asn Asn Asn Glu Cys Glu Ser Asn Pro Cys Val Asn Gly Gly Thr Cys
755 760 765

aaa gac atg acc agt ggc atc gtg tgc acc tgc cgg gag ggc ttc agc 2352
Lys Asp Met Thr Ser Gly Ile Val Cys Thr Cys Arg Glu Gly Phe Ser
770 775 780

ggt ccc aac tgc cag acc aac atc aac gag tgt gcg tcc aac cca tgt 2400
Gly Pro Asn Cys Gln Thr Asn Ile Asn Glu Cys Ala Ser Asn Pro Cys
785 790 795 800

ctg aac aag ggc acg tgt att gac gac gtt gcc ggg tac aag tgc aac 2448
Leu Asn Lys Gly Thr Cys Ile Asp Asp Val Ala Gly Tyr Lys Cys Asn
805 810 815

tgc ctg ctg ccc tac aca ggt gcc acg tgt gag gtg gtg ctg gcc ccg 2496
Cys Leu Leu Pro Tyr Thr Gly Ala Thr Cys Glu Val Val Leu Ala Pro

820 825 830

tgt gcc ccc agc ccc tgc aga aac ggc ggg gag tgc agg caa tcc gag 2544
Cys Ala Pro Ser Pro Cys Arg Asn Gly Gly Glu Cys Arg Gln Ser Glu
835 840 845

gac tat gag agc ttc tcc tgt gtc tgc ccc acg gct ggg gcc aaa ggg 2592
Asp Tyr Glu Ser Phe Ser Cys Val Cys Pro Thr Ala Gly Ala Lys Gly
850 855 860

cag acc tgt gag gtc gac atc aac gag tgc gtt ctg agc ccg tgc cgg 2640
Gln Thr Cys Glu Val Asp Ile Asn Glu Cys Val Leu Ser Pro Cys Arg
865 870 875 880

cac ggc gca tcc tgc cag aac acc cac ggc ggc tac cgc tgc cac tgc 2688
His Gly Ala Ser Cys Gln Asn Thr His Gly Gly Tyr Arg Cys His Cys
885 890 895

cag gcc ggc tac agt ggg cgc aac tgc gag acc gac atc gac gac tgc 2736
Gln Ala Gly Tyr Ser Gly Arg Asn Cys Glu Thr Asp Ile Asp Asp Cys
900 905 910

cgg ccc aac ccg tgt cac aac ggg ggc tcc tgc aca gac ggc atc aac 2784
Arg Pro Asn Pro Cys His Asn Gly Gly Ser Cys Thr Asp Gly Ile Asn
915 920 925

acg gcc ttc tgc gac tgc ctg ccc ggc ttc cgg ggc act ttc tgt gag 2832
Thr Ala Phe Cys Asp Cys Leu Pro Gly Phe Arg Gly Thr Phe Cys Glu
930 935 940

gag gac atc aac gag tgt gcc agt gac ccc tgc cgc aac ggg gcc aac 2880
Glu Asp Ile Asn Glu Cys Ala Ser Asp Pro Cys Arg Asn Gly Ala Asn
945 950 955 960

tgc acg gac tgc gtg gac agc tac acg tgc acc tgc ccc gca ggc ttc 2928
Cys Thr Asp Cys Val Asp Ser Tyr Thr Cys Thr Cys Pro Ala Gly Phe
965 970 975

agc ggg atc cac tgt gag aac aac acg cct gac tgc aca gag agc tcc 2976
Ser Gly Ile His Cys Glu Asn Asn Thr Pro Asp Cys Thr Glu Ser Ser
980 985 990

tgc ttc aac ggt ggc acc tgc gtg gac ggc atc aac tcg ttc acc tgc 3024
Cys Phe Asn Gly Gly Thr Cys Val Asp Gly Ile Asn Ser Phe Thr Cys
995 1000 1005

ctg tgt cca ccc ggc ttc acg ggc agc tac tgc cag cac gta gtc aat 3072
Leu Cys Pro Pro Gly Phe Thr Gly Ser Tyr Cys Gln His Val Val Asn
1010 1015 1020

gag tgc gac tca cga ccc tgc ctg cta ggc ggc acc tgt cag gac ggt 3120
Glu Cys Asp Ser Arg Pro Cys Leu Leu Gly Gly Thr Cys Gln Asp Gly
1025 1030 1035 1040

cgc ggt ctc cac agg tgc acc tgc ccc cag ggc tac act ggc ccc aac 3168
Arg Gly Leu His Arg Cys Thr Cys Pro Gln Gly Tyr Thr Gly Pro Asn
1045 1050 1055

tgc cag aac ctt gtg cac tgg tgt gac tcc tcg ccc tgc aag aac ggc 3216
Cys Gln Asn Leu Val His Trp Cys Asp Ser Ser Pro Cys Lys Asn Gly
1060 1065 1070

ggc aaa tgc tgg cag acc cac acc cag tac cgc tgc gag tgc ccc agc 3264
Gly Lys Cys Trp Gln Thr His Thr Gln Tyr Arg Cys Glu Cys Pro Ser
1075 1080 1085

ggc tgg acc ggc ctt tac tgc gac gtg ccc agc gtg tcc tgt gag gtg 3312

Gly Trp Thr Gly Leu Tyr Cys Asp Val Pro Ser Val Ser Cys Glu
Val
1090 1095 1100

gct gcg cag cga caa ggt gtt gac gtt gcc cgc ctg tgc cag cat
gga 3360
Ala Ala Gln Arg Gln Gly Val Asp Val Ala Arg Leu Cys Gln His
Gly
1105 1110 1115
1120

ggg ctc tgt gtg gac gcg ggc aac acg cac cac tgc cgc tgc cag
gcg 3408
Gly Leu Cys Val Asp Ala Gly Asn Thr His His Cys Arg Cys Gln
Ala
1125 1130 1135

ggc tac aca ggc agc tac tgt gag gac ctg gtg gac gag tgc tca
ccc 3456
Gly Tyr Thr Gly Ser Tyr Cys Glu Asp Leu Val Asp Glu Cys Ser
Pro
1140 1145 1150

agc ccc tgc cag aac ggg gcc acc tgc acg gac tac ctg ggc ggc
tac 3504
Ser Pro Cys Gln Asn Gly Ala Thr Cys Thr Asp Tyr Leu Gly Gly
Tyr
1155 1160 1165

tcc tgc aag tgc gtg gcc ggc tac cac ggg gtg aac tgc tct gag
gag 3552
Ser Cys Lys Cys Val Ala Gly Tyr His Gly Val Asn Cys Ser Glu
Glu
1170 1175 1180

atc gac gag tgc ctc tcc cac ccc tgc cag aac ggg ggc acc tgc
ctc 3600
Ile Asp Glu Cys Leu Ser His Pro Cys Gln Asn Gly Gly Thr Cys
Leu
1185 1190 1195
1200

gac ctc ccc aac acc tac aag tgc tcc tgc cca cgg ggc act cag
ggt 3648
Asp Leu Pro Asn Thr Tyr Lys Cys Ser Cys Pro Arg Gly Thr Gln
Gly
1205 1210 1215

gtg cac tgt gag atc aac gtg gac gac tgc aat ccc ccc gtt gac
ccc 3696
Val His Cys Glu Ile Asn Val Asp Asp Cys Asn Pro Pro Val Asp
Pro

1220 1225 1230

gtg tcc cgg agc ccc aag tgc ttt aac aac ggc acc tgc gtg gac
cag 3744
Val Ser Arg Ser Pro Lys Cys Phe Asn Asn Gly Thr Cys Val Asp
Gln
1235 1240 1245

gtg ggc ggc tac agc tgc acc tgc ccg ccg ggc ttc gtg ggt gag
cgc 3792
Val Gly Gly Tyr Ser Cys Thr Cys Pro Pro Gly Phe Val Gly Glu
Arg
1250 1255 1260

tgt gag ggg gat gtc aac gag tgc ctg tcc aat ccc tgc gac gcc
cgt 3840
Cys Glu Gly Asp Val Asn Glu Cys Leu Ser Asn Pro Cys Asp Ala
Arg
1265 1270 1275
1280

ggc acc cag aac tgc gtg cag cgc gtc aat gac ttc cac tgc gag
tgc 3888
Gly Thr Gln Asn Cys Val Gln Arg Val Asn Asp Phe His Cys Glu
Cys
1285 1290 1295

cgt gct ggt cac acc ggg cgc cgc tgc gag tcc gtc atc aat ggc
tgc 3936
Arg Ala Gly His Thr Gly Arg Arg Cys Glu Ser Val Ile Asn Gly
Cys
1300 1305 1310

aaa ggc aag ccc tgc aag aat ggg ggc acc tgc gcc gtg gcc tcc
aac 3984
Lys Gly Lys Pro Cys Lys Asn Gly Gly Thr Cys Ala Val Ala Ser
Asn
1315 1320 1325

acc gcc cgc ggg ttc atc tgc aag tgc cct gcg ggc ttc gag ggc
gcc 4032
Thr Ala Arg Gly Phe Ile Cys Lys Cys Pro Ala Gly Phe Glu Gly
Ala
1330 1335 1340

acg tgt gag aat gac gct cgt acc tgc ggc agc ctg cgc tgc ctc
aac 4080
Thr Cys Glu Asn Asp Ala Arg Thr Cys Gly Ser Leu Arg Cys Leu
Asn
1345 1350 1355
1360

ggc ggc aca tgc atc tcc ggc ccg cgc agc ccc acc tgc ctg tgc ctg 4128
Gly Gly Thr Cys Ile Ser Gly Pro Arg Ser Pro Thr Cys Leu Cys Leu
1365 1370 1375

ggc ccc ttc acg ggc ccc gaa tgc cag ttc ccg gcc agc agc ccc tgc 4176
Gly Pro Phe Thr Gly Pro Glu Cys Gln Phe Pro Ala Ser Ser Pro Cys
1380 1385 1390

ctg ggc ggc aac ccc tgc tac aac cag ggg acc tgt gag ccc aca tcc 4224
Leu Gly Gly Asn Pro Cys Tyr Asn Gln Gly Thr Cys Glu Pro Thr Ser
1395 1400 1405

gag agc ccc ttc tac cgt tgc ctg tgc ccc gcc aaa ttc aac ggg ctc 4272
Glu Ser Pro Phe Tyr Arg Cys Leu Cys Pro Ala Lys Phe Asn Gly Leu
1410 1415 1420

ttg tgc cac atc ctg gac tac agc ttc ggg ggt ggg gcc ggg cgc gac 4320
Leu Cys His Ile Leu Asp Tyr Ser Phe Gly Gly Gly Ala Gly Arg Asp
1425 1430 1435 1440

atc ccc ccg ccg ctg atc gag gag gcg tgc gag ctg ccc gag tgc cag 4368
Ile Pro Pro Pro Leu Ile Glu Glu Ala Cys Glu Leu Pro Glu Cys Gln
1445 1450 1455

gag gac gcg ggc aac aag gtc tgc agc ctg cag tgc aac aac cac gcg 4416
Glu Asp Ala Gly Asn Lys Val Cys Ser Leu Gln Cys Asn Asn His Ala
1460 1465 1470

tgc ggc tgg gac ggc ggt gac tgc tcc ctc aac ttc aat gac ccc tgg 4464
Cys Gly Trp Asp Gly Gly Asp Cys Ser Leu Asn Phe Asn Asp Pro Trp
1475 1480 1485

aag aac tgc acg cag tct ctg cag tgc tgg aag tac ttc agt gac ggc 4512

Lys Asn Cys Thr Gln Ser Leu Gln Cys Trp Lys Tyr Phe Ser Asp Gly
1490 1495 1500

cac tgt gac agc cag tgc aac tca gcc ggc tgc ctc ttc gac ggc ttt 4560
His Cys Asp Ser Gln Cys Asn Ser Ala Gly Cys Leu Phe Asp Gly Phe
1505 1510 1515 1520

gac tgc cag cgt gcg gaa ggc cag tgc aac ccc ctg tac gac cag tac 4608
Asp Cys Gln Arg Ala Glu Gly Gln Cys Asn Pro Leu Tyr Asp Gln Tyr
1525 1530 1535

tgc aag gac cac ttc agc gac ggg cac tgc gac cag ggc tgc aac agc 4656
Cys Lys Asp His Phe Ser Asp Gly His Cys Asp Gln Gly Cys Asn Ser
1540 1545 1550

gcg gag tgc gag tgg gac ggg ctg gac tgt gcg gag cat gta ccc gag 4704
Ala Glu Cys Glu Trp Asp Gly Leu Asp Cys Ala Glu His Val Pro Glu
1555 1560 1565

agg ctg gcg gcc ggc acg ctg gtg gtg gtg gtg ctg atg ccg ccg gag 4752
Arg Leu Ala Ala Gly Thr Leu Val Val Val Val Leu Met Pro Pro Glu
1570 1575 1580

cag ctg cgc aac agc tcc ttc cac ttc ctg cgg gag ctc agc cgc gtg 4800
Gln Leu Arg Asn Ser Ser Phe His Phe Leu Arg Glu Leu Ser Arg Val
1585 1590 1595 1600

ctg cac acc aac gtg gtc ttc aag cgt gac gca cac ggc cag cag atg 4848
Leu His Thr Asn Val Val Phe Lys Arg Asp Ala His Gly Gln Gln Met
1605 1610 1615

atc ttc ccc tac tac ggc cgc gag gag gag ctg cgc aag cac ccc atc 4896
Ile Phe Pro Tyr Tyr Gly Arg Glu Glu Glu Leu Arg Lys His Pro Ile

1620 1625 1630

aag cgt gcc gcc gag ggc tgg gcc gca cct gac gcc ctg ctg ggc cag 4944
Lys Arg Ala Ala Glu Gly Trp Ala Ala Pro Asp Ala Leu Leu Gly Gln
1635 1640 1645

gtg aag gcc tcg ctg ctc cct ggt ggc agc gag ggt ggg cgg cgg cgg 4992
Val Lys Ala Ser Leu Leu Pro Gly Gly Ser Glu Gly Gly Arg Arg Arg
1650 1655 1660

agg gag ctg gac ccc atg gac gtc cgc ggc tcc atc gtc tac ctg gag 5040
Arg Glu Leu Asp Pro Met Asp Val Arg Gly Ser Ile Val Tyr Leu Glu
1665 1670 1675 1680

att gac aac cgg cag tgt gtg cag gcc tcc tcg cag tgc ttc cag agt 5088
Ile Asp Asn Arg Gln Cys Val Gln Ala Ser Ser Gln Cys Phe Gln Ser
1685 1690 1695

gcc acc gat gtg gcc gca ttc ctg gga gcg ctc gcc tcg ctg ggc agc 5136
Ala Thr Asp Val Ala Ala Phe Leu Gly Ala Leu Ala Ser Leu Gly Ser
1700 1705 1710

ctc aac atc ccc tac aag atc gag gcc gtg cag agt gag acc gtg gag 5184
Leu Asn Ile Pro Tyr Lys Ile Glu Ala Val Gln Ser Glu Thr Val Glu
1715 1720 1725

ccg ccc ccg ccg gcg cag ctg cac ttc atg tac gtg gcg gcg gcc gcc 5232
Pro Pro Pro Pro Ala Gln Leu His Phe Met Tyr Val Ala Ala Ala Ala
1730 1735 1740

ttt gtg ctt ctg ttc ttc gtg ggc tgc ggg gtg ctg ctg tcc cgc aag 5280
Phe Val Leu Leu Phe Phe Val Gly Cys Gly Val Leu Leu Ser Arg Lys
1745 1750 1755 1760

cgc cgg cgg cag cat ggc cag ctc tgg ttc cct gag ggc ttc aaa gtg 5328
Arg Arg Arg Gln His Gly Gln Leu Trp Phe Pro Glu Gly Phe Lys Val
1765 1770 1775

tct gag gcc agc aag aag aag cgg cgg gag ccc ctc ggc gag gac tcc 5376
Ser Glu Ala Ser Lys Lys Lys Arg Arg Glu Pro Leu Gly Glu Asp Ser
1780 1785 1790

gtg ggc ctc aag ccc ctg aag aac gct tca gac ggt gcc ctc atg gac 5424
Val Gly Leu Lys Pro Leu Lys Asn Ala Ser Asp Gly Ala Leu Met Asp
1795 1800 1805

gac aac cag aat gag tgg ggg gac gag gac ctg gag acc aag aag ttc 5472
Asp Asn Gln Asn Glu Trp Gly Asp Glu Asp Leu Glu Thr Lys Lys Phe
1810 1815 1820

cgg ttc gag gag ccc gtg gtt ctg cct gac ctg gac gac cag aca gac 5520
Arg Phe Glu Glu Pro Val Val Leu Pro Asp Leu Asp Asp Gln Thr Asp
1825 1830 1835
1840

cac cgg cag tgg act cag cag cac ctg gat gcc gct gac ctg cgc atg 5568
His Arg Gln Trp Thr Gln Gln His Leu Asp Ala Ala Asp Leu Arg Met
1845 1850 1855

tct gcc atg gcc ccc aca ccg ccc cag ggt gag gtt gac gcc gac tgc 5616
Ser Ala Met Ala Pro Thr Pro Pro Gln Gly Glu Val Asp Ala Asp Cys
1860 1865 1870

atg gac gtc aat gtc cgc ggg cct gat ggc ttc acc ccg ctc atg atc 5664
Met Asp Val Asn Val Arg Gly Pro Asp Gly Phe Thr Pro Leu Met Ile
1875 1880 1885

gcc tcc tgc agc ggg ggc ggc ctg gag acg ggc aac agc gag gaa gag 5712

Ala Ser Cys Ser Gly Gly Gly Leu Glu Thr Gly Asn Ser Glu Glu
Glu
1890 1895 1900

gag gac gcg ccg gcc gtc atc tcc gac ttc atc tac cag ggc gcc
agc 5760
Glu Asp Ala Pro Ala Val Ile Ser Asp Phe Ile Tyr Gln Gly Ala
Ser
1905 1910 1915
1920

ctg cac aac cag aca gac cgc acg ggc gag acc gcc ttg cac ctg
gcc 5808
Leu His Asn Gln Thr Asp Arg Thr Gly Glu Thr Ala Leu His Leu
Ala
1925 1930 1935

gcc cgc tac tca cgc tct gat gcc gcc aag cgc ctg ctg gag gcc
agc 5856
Ala Arg Tyr Ser Arg Ser Asp Ala Ala Lys Arg Leu Leu Glu Ala
Ser
1940 1945 1950

gca gat gcc aac atc cag gac aac atg ggc cgc acc ccg ctg cat
gcg 5904
Ala Asp Ala Asn Ile Gln Asp Asn Met Gly Arg Thr Pro Leu His
Ala
1955 1960 1965

gct gtg tct gcc gac gca caa ggt gtc ttc cag atc ctg atc cgg
aac 5952
Ala Val Ser Ala Asp Ala Gln Gly Val Phe Gln Ile Leu Ile Arg
Asn
1970 1975 1980

cga gcc aca gac ctg gat gcc cgc atg cat gat ggc acg acg cca
ctg 6000
Arg Ala Thr Asp Leu Asp Ala Arg Met His Asp Gly Thr Thr Pro
Leu
1985 1990 1995
2000

atc ctg gct gcc cgc ctg gcc gtg gag ggc atg ctg gag gac ctc
atc 6048
Ile Leu Ala Ala Arg Leu Ala Val Glu Gly Met Leu Glu Asp Leu
Ile
2005 2010 2015

aac tca cac gcc gac gtc aac gcc gta gat gac ctg ggc aag tcc
gcc 6096
Asn Ser His Ala Asp Val Asn Ala Val Asp Asp Leu Gly Lys Ser
Ala

2020 2025 2030

ctg cac tgg gcc gcc gcc gtg aac aat gtg gat gcc gca gtt gtg ctc 6144
Leu His Trp Ala Ala Ala Val Asn Asn Val Asp Ala Ala Val Val Leu
2035 2040 2045

ctg aag aac ggg gct aac aaa gat atg cag aac aac agg gag gag aca 6192
Leu Lys Asn Gly Ala Asn Lys Asp Met Gln Asn Asn Arg Glu Glu Thr
2050 2055 2060

ccc ctg ttt ctg gcc gcc cgg gag ggc agc tac gag acc gcc aag gtg 6240
Pro Leu Phe Leu Ala Ala Arg Glu Gly Ser Tyr Glu Thr Ala Lys Val
2065 2070 2075
2080

ctg ctg gac cac ttt gcc aac cgg gac atc acg gat cat atg gac cgc 6288
Leu Leu Asp His Phe Ala Asn Arg Asp Ile Thr Asp His Met Asp Arg
2085 2090 2095

ctg ccg cgc gac atc gca cag gag cgc atg cat cac gac atc gtg agg 6336
Leu Pro Arg Asp Ile Ala Gln Glu Arg Met His His Asp Ile Val Arg
2100 2105 2110

ctg ctg gac gag tac aac ctg gtg cgc agc ccg cag ctg cac gga gcc 6384
Leu Leu Asp Glu Tyr Asn Leu Val Arg Ser Pro Gln Leu His Gly Ala
2115 2120 2125

ccg ctg ggg ggc acg ccc acc ctg tcg ccc ccg ctc tgc tcg ccc aac 6432
Pro Leu Gly Gly Thr Pro Thr Leu Ser Pro Pro Leu Cys Ser Pro Asn
2130 2135 2140

ggc tac ctg ggc agc ctc aag ccc ggc gtg cag ggc aag aag gtc cgc 6480
Gly Tyr Leu Gly Ser Leu Lys Pro Gly Val Gln Gly Lys Lys Val Arg
2145 2150 2155
2160

aag ccc agc agc aaa ggc ctg gcc tgt gga agc aag gag gcc aag gac 6528
Lys Pro Ser Ser Lys Gly Leu Ala Cys Gly Ser Lys Glu Ala Lys Asp
2165 2170 2175

ctc aag gca cgg agg aag aag tcc cag gat ggc aag ggc tgc ctg ctg 6576
Leu Lys Ala Arg Arg Lys Lys Ser Gln Asp Gly Lys Gly Cys Leu Leu
2180 2185 2190

gac agc tcc ggc atg ctc tcg ccc gtg gac tcc ctg gag tca ccc cat 6624
Asp Ser Ser Gly Met Leu Ser Pro Val Asp Ser Leu Glu Ser Pro His
2195 2200 2205

ggc tac ctg tca gac gtg gcc tcg ccg cca ctg ctg ccc tcc ccg ttc 6672
Gly Tyr Leu Ser Asp Val Ala Ser Pro Pro Leu Leu Pro Ser Pro Phe
2210 2215 2220

cag cag tct ccg tcc gtg ccc ctc aac cac ctg cct ggg atg ccc gac 6720
Gln Gln Ser Pro Ser Val Pro Leu Asn His Leu Pro Gly Met Pro Asp
2225 2230 2235
2240

acc cac ctg ggc atc ggg cac ctg aac gtg gcg gcc aag ccc gag atg 6768
Thr His Leu Gly Ile Gly His Leu Asn Val Ala Ala Lys Pro Glu Met
2245 2250 2255

gcg gcg ctg ggt ggg ggc ggc cgg ctg gcc ttt gag act ggc cca cct 6816
Ala Ala Leu Gly Gly Gly Gly Arg Leu Ala Phe Glu Thr Gly Pro Pro
2260 2265 2270

cgt ctc tcc cac ctg cct gtg gcc tct ggc acc agc acc gtc ctg ggc 6864
Arg Leu Ser His Leu Pro Val Ala Ser Gly Thr Ser Thr Val Leu Gly
2275 2280 2285

tcc agc agc gga ggg gcc ctg aat ttc act gtg ggc ggg tcc acc agt 6912

Ser Ser Ser Gly Gly Ala Leu Asn Phe Thr Val Gly Gly Ser Thr Ser
2290 2295 2300

ttg aat ggt caa tgc gag tgg ctg tcc cgg ctg cag agc ggc atg gtg 6960
Leu Asn Gly Gln Cys Glu Trp Leu Ser Arg Leu Gln Ser Gly Met Val
2305 2310 2315 2320

ccg aac caa tac aac cct ctg cgg ggg agt gtg gca cca ggc ccc ctg 7008
Pro Asn Gln Tyr Asn Pro Leu Arg Gly Ser Val Ala Pro Gly Pro Leu
2325 2330 2335

agc aca cag gcc ccc tcc ctg cag cat ggc atg gta ggc ccg ctg cac 7056
Ser Thr Gln Ala Pro Ser Leu Gln His Gly Met Val Gly Pro Leu His
2340 2345 2350

agt agc ctt gct gcc agc gcc ctg tcc cag atg atg agc tac cag ggc 7104
Ser Ser Leu Ala Ala Ser Ala Leu Ser Gln Met Met Ser Tyr Gln Gly
2355 2360 2365

ctg ccc agc acc cgg ctg gcc acc cag cct cac ctg gtg cag acc cag 7152
Leu Pro Ser Thr Arg Leu Ala Thr Gln Pro His Leu Val Gln Thr Gln
2370 2375 2380

cag gtg cag cca caa aac tta cag atg cag cag cag aac ctg cag cca 7200
Gln Val Gln Pro Gln Asn Leu Gln Met Gln Gln Gln Asn Leu Gln Pro
2385 2390 2395 2400

gca aac atc cag cag cag caa agc ctg cag ccg cca cca cca cca cca 7248
Ala Asn Ile Gln Gln Gln Gln Ser Leu Gln Pro Pro Pro Pro Pro Pro
2405 2410 2415

cag ccg cac ctt ggc gtg agc tca gca gcc agc ggc cac ctg ggc cgg 7296
Gln Pro His Leu Gly Val Ser Ser Ala Ala Ser Gly His Leu Gly Arg

2420 2425 2430

agc ttc ctg agt gga gag ccg agc cag gca gac gtg cag cca ctg ggc 7344
Ser Phe Leu Ser Gly Glu Pro Ser Gln Ala Asp Val Gln Pro Leu Gly
2435 2440 2445

ccc agc agc ctg gcg gtg cac act att ctg ccc cag gag agc ccc gcc 7392
Pro Ser Ser Leu Ala Val His Thr Ile Leu Pro Gln Glu Ser Pro Ala
2450 2455 2460

ctg ccc acg tcg ctg cca tcc tcg ctg gtc cca ccc gtg acc gca gcc 7440
Leu Pro Thr Ser Leu Pro Ser Ser Leu Val Pro Pro Val Thr Ala Ala
2465 2470 2475
2480

cag ttc ctg acg ccc ccc tcg cag cac agc tac tcc tcg cct gtg gac 7488
Gln Phe Leu Thr Pro Pro Ser Gln His Ser Tyr Ser Ser Pro Val Asp
2485 2490 2495

aac acc ccc agc cac cag cta cag gtg cct gag cac ccc ttc ctc acc 7536
Asn Thr Pro Ser His Gln Leu Gln Val Pro Glu His Pro Phe Leu Thr
2500 2505 2510

cct tcg ccg gag tcg ccc gac caa tgg tcg tcc tcg tcg ccg cac tct 7584
Pro Ser Pro Glu Ser Pro Asp Gln Trp Ser Ser Ser Ser Pro His Ser
2515 2520 2525

aat gtg tct gac tgg tct gag ggc gtg tcg tcg ccc ccg acc tcc atg 7632
Asn Val Ser Asp Trp Ser Glu Gly Val Ser Ser Pro Pro Thr Ser Met
2530 2535 2540

cag tcc cag atc gcg cgc atc ccg gag gcg ttc aag taa ta 7673
Gln Ser Gln Ile Ala Arg Ile Pro Glu Ala Phe Lys
2545 2550 2555

<210> 2
<211> 3228
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (716)..(2689)

<400> 2
tcctccactc tgggaacagc tcacaccccc aaagtggtga ccaccatggc cactatgccc 60

acagccactg cctccacggt tcccagctcg tccaccgtgg ggaccacccg cacccctgca 120

gtgctcccca gcagcctgcc aaccttcagc gtgtccactg tgtcctcctc agtcctcacc 180

accctgagac ccactggctt ccccagctcc cacttttcta ctccctgctt ctgcagggca 240

tttggacagt ttttctcgcc cggggaagtc atctacaata agaccgaccg agccggctgc 300

catttctacg cagtgtgcaa tcagcactgt gacattgacc gcttccaggg cgcctgtccc 360

acctccccac cgccagtgtc ctccgccccg ctgtcctcgc cctcccctgc ccctggctgt 420

gacaatgcca tccctctccg gcaggtgaat gagacctgga ccctggagaa ctgcacggtg 480

gccaggtgcg tgggtgacaa ccgtgtcgtc ctgctggacc caaagcctgt ggccaacgtc 540

acctgcgtga acaagcacct gcccatcaaa gtgtcggacc cgagccagcc ctgtgacttc 600

cactatgagt gcgagtgcat ctgcagcatg tggggcggct cccactattc cacctttgac 660

ggcacctctt acaccttccg ggcaactgc acctatgtcc tcatgagaga gatcc atg 718

Met
1

cac gct ttg gga atc tca gcc tct acc tgg aca acc act act gca cgg 766
His Ala Leu Gly Ile Ser Ala Ser Thr Trp Thr Thr Thr Thr Ala Arg
5 10 15

cct ctg cca ctg ccg ctg ccg ccc gct gcc ccc gcg gcc ctc agc atc 814
Pro Leu Pro Leu Pro Leu Pro Pro Ala Ala Pro Ala Ala Leu Ser Ile
20 25 30

cac tac aag tcc atg gat atc gtc ctc act gtc acc atg gtg cat ggg 862
His Tyr Lys Ser Met Asp Ile Val Leu Thr Val Thr Met Val His Gly
35 40 45

aag gag gag ggc ctg atc ctg ttt gac cca att ccg gtg agc agc ggt 910
Lys Glu Glu Gly Leu Ile Leu Phe Asp Pro Ile Pro Val Ser Ser Gly
50 55 60 65

ttc agc aag aac ggc gtg ctt gtg tct gtg ctg ggg acc acc acc atg 958
Phe Ser Lys Asn Gly Val Leu Val Ser Val Leu Gly Thr Thr Thr Met
70 75 80

gct gtg gac att cct gcc ctg ggc gtg agc gtc acc ttc aat ggc caa 1006
Ala Val Asp Ile Pro Ala Leu Gly Val Ser Val Thr Phe Asn Gly Gln
85 90 95

gtg ttc cag gcc cgg ctg ccc tac agc ctc ttc cac aac aac acc gag 1054
Val Phe Gln Ala Arg Leu Pro Tyr Ser Leu Phe His Asn Asn Thr Glu
100 105 110

ggc cag tgc ggc acc tgc acc aac aac cag agg gac gac tgt ctc cag 1102
Gly Gln Cys Gly Thr Cys Thr Asn Asn Gln Arg Asp Asp Cys Leu Gln
115 120 125

cgg gac gga acc act gcc gcc agt tgc aag gac atg gcc aag acg tgg 1150

Arg Asp Gly Thr Thr Ala Ala Ser Cys Lys Asp Met Ala Lys Thr
Trp
130 135 140
145

ctg gtc ccc gac agc aga aag gat ggc tgt tgg gcc ccg act ggc
aca 1198
Leu Val Pro Asp Ser Arg Lys Asp Gly Cys Trp Ala Pro Thr Gly
Thr
150 155 160

ccc ccc act gcc agc ccc gca gcc ccg gtg tct agc aca ccc acc
ccc 1246
Pro Pro Thr Ala Ser Pro Ala Ala Pro Val Ser Ser Thr Pro Thr
Pro
165 170 175

acc cca tgc cca cca cag ctg ctc tgt gat ctg atg ctg agc cag
gtc 1294
Thr Pro Cys Pro Pro Gln Leu Leu Cys Asp Leu Met Leu Ser Gln
Val
180 185 190

ttt gct gag tgc cac acc ctt ctg ccc ccg ggc cca ttc ttc aac
gcc 1342
Phe Ala Glu Cys His Thr Leu Leu Pro Pro Gly Pro Phe Phe Asn
Ala
195 200 205

tgc atc agc gac cac tgc agg ggc cgc ctt gag gtg ccc tgc cag
agc 1390
Cys Ile Ser Asp His Cys Arg Gly Arg Leu Glu Val Pro Cys Gln
Ser
210 215 220
225

ctg gag gct tac gca gag ctc tgc cgc gcc cgg gga gtg tgc agt
gac 1438
Leu Glu Ala Tyr Ala Glu Leu Cys Arg Ala Arg Gly Val Cys Ser
Asp
230 235 240

tgg cga ggt gca acc ggt ggc ctg tgc gac ctc acc tgc cca ccc
acc 1486
Trp Arg Gly Ala Thr Gly Gly Leu Cys Asp Leu Thr Cys Pro Pro
Thr
245 250 255

aaa gtg tac aag cca tgc ggc ccc ata cag cct gcc acc tgc aac
tct 1534
Lys Val Tyr Lys Pro Cys Gly Pro Ile Gln Pro Ala Thr Cys Asn
Ser

260 265 270

agg aac cag agc cca cag ctg gag ggg atg gcg gag ggc tgc ttc tgc 1582
Arg Asn Gln Ser Pro Gln Leu Glu Gly Met Ala Glu Gly Cys Phe Cys
275 280 285

cct gag gac cag atc ctc ttc aac gca cac atg ggc atc tgc gtg cag 1630
Pro Glu Asp Gln Ile Leu Phe Asn Ala His Met Gly Ile Cys Val Gln
290 295 300 305

gcc tgc ccc tgc gtg gga ccc gat ggg ttt cct aaa ttt ccc ggg gag 1678
Ala Cys Pro Cys Val Gly Pro Asp Gly Phe Pro Lys Phe Pro Gly Glu
310 315 320

cgg tgg gtc agc aac tgc cag tcc tgc gtg tgt gac gag ggt tca gtg 1726
Arg Trp Val Ser Asn Cys Gln Ser Cys Val Cys Asp Glu Gly Ser Val
325 330 335

tcg gtg cag tgc aag ccc ctg ccc tgt gac gcc cag ggt cag ccc ccg 1774
Ser Val Gln Cys Lys Pro Leu Pro Cys Asp Ala Gln Gly Gln Pro Pro
340 345 350

ccg tgc aac cgt ccc ggc ttc gta acc gtg acc agg ccc cgg gcc gag 1822
Pro Cys Asn Arg Pro Gly Phe Val Thr Val Thr Arg Pro Arg Ala Glu
355 360 365

aac ccc tgc tgc ccc gag acg gtg tgc gtg tgc aac aca acc acc tgc 1870
Asn Pro Cys Cys Pro Glu Thr Val Cys Val Cys Asn Thr Thr Thr Cys
370 375 380 385

ccc cag agc ctg cct gtg tgc ccg cca ggg cag gag tcc atc tgc acc 1918
Pro Gln Ser Leu Pro Val Cys Pro Pro Gly Gln Glu Ser Ile Cys Thr
390 395 400

cag gag gag ggc gac tgc tgt ccc acc ttc cgc tgc aga cct cag ctg 1966
Gln Glu Glu Gly Asp Cys Cys Pro Thr Phe Arg Cys Arg Pro Gln Leu
405 410 415

tgt tcg tac aat ggc acc ttc tac ggg gtt ggt gca acc ttc cca ggc 2014
Cys Ser Tyr Asn Gly Thr Phe Tyr Gly Val Gly Ala Thr Phe Pro Gly
420 425 430

gcc ctt ccc tgc cac atg tgt acc tgc ctc tct ggg gac acc cag gac 2062
Ala Leu Pro Cys His Met Cys Thr Cys Leu Ser Gly Asp Thr Gln Asp
435 440 445

cca acg gtg caa tgt cag gag gat gcc tgc aac aat act acc tgt ccc 2110
Pro Thr Val Gln Cys Gln Glu Asp Ala Cys Asn Asn Thr Thr Cys Pro
450 455 460 465

cag ggc ttt gag tac aag aga gtg gcc ggg cag tgc tgt ggg gag tgc 2158
Gln Gly Phe Glu Tyr Lys Arg Val Ala Gly Gln Cys Cys Gly Glu Cys
470 475 480

gtc cag acc gcc tgc ctc acg ccc gat ggc cag cca gtc cag ctg aat 2206
Val Gln Thr Ala Cys Leu Thr Pro Asp Gly Gln Pro Val Gln Leu Asn
485 490 495

gaa acc tgg gtc aac agc cat gtg gac aac tgc acc gtg tac ctc tgt 2254
Glu Thr Trp Val Asn Ser His Val Asp Asn Cys Thr Val Tyr Leu Cys
500 505 510

gag gct gag ggt gga gtc cat ttg ctg acc cca cag cct gca tcc tgc 2302
Glu Ala Glu Gly Gly Val His Leu Leu Thr Pro Gln Pro Ala Ser Cys
515 520 525

cca gat gtg tcc agc tgc agg ggg agc ctc agg aaa acc ggc tgc tgc 2350

```
 Pro Asp Val Ser Ser Cys Arg Gly Ser Leu Arg Lys Thr Gly Cys
Cys
 530                 535                 540
545

 tac tcc tgt gag gag gac tcc tgt caa gtc cgc atc aac acg acc
atc     2398
 Tyr Ser Cys Glu Glu Asp Ser Cys Gln Val Arg Ile Asn Thr Thr
Ile
                 550                 555                 560

 ctg tgg cac cag ggc tgc gag acc gag gtc aac atc acc ttc tgc
gag     2446
 Leu Trp His Gln Gly Cys Glu Thr Glu Val Asn Ile Thr Phe Cys
Glu
             565                 570                 575

 ggc tcc tgc ccc gga gcg tcc aag tac tca gca gag gcc cag gcc
atg     2494
 Gly Ser Cys Pro Gly Ala Ser Lys Tyr Ser Ala Glu Ala Gln Ala
Met
         580                 585                 590

 cag cac cag tgc acc tgc tgc cag gag agg cgg gtc cac gag gag
acg     2542
 Gln His Gln Cys Thr Cys Cys Gln Glu Arg Arg Val His Glu Glu
Thr
     595                 600                 605

 gtg ccc ttg cac tgt cct aac ggc tca gcc atc ctg cac acc tac
acc     2590
 Val Pro Leu His Cys Pro Asn Gly Ser Ala Ile Leu His Thr Tyr
Thr
 610                 615                 620
625

 cac gtg gat gag tgt ggc tgc acg ccc ttc tgt gtc cct gcg ccc
atg     2638
 His Val Asp Glu Cys Gly Cys Thr Pro Phe Cys Val Pro Ala Pro
Met
                 630                 635                 640

 gct ccc cca cac acc cgt ggc ttc ccg gcc cag gag gcc act gct
gtc     2686
 Ala Pro Pro His Thr Arg Gly Phe Pro Ala Gln Glu Ala Thr Ala
Val
             645                 650                 655

 tga gaac gttctgcctc catccccatg ctctgtccac ctggagccag
gatgtgcatt     2743
```

gtctgatcat gaaaaccttg ggcctcctct gcggagcccc ccggcctgtg tgtggcaccc 2803

cgcgctccgt gctcctgctg cccaccccgt gggtgaaacc ggccccagaa gggtgagggg 2863

ccagcaggac cctttcggga gggcgccact caggagtcct accctgggag agcctgtggc 2923

ccaccttggc cttgcccctc cctgatgtca ctgggacgcc ctggaacaaa ctaagcatgt 2983

gcgggcctat gtgtccctgc cacggccgga cgcccgcgca gcacggattc cagctggcca 3043

cgtccggccg ctggggcaga caggctggtc caggcaaggc cagctgctgc caggaagctg 3103

cgacaggcaa gcggccgcct gtccatgcct gctgcagggt aactcagggc tgaggtcgca 3163

acggccaggt cagagagggg tcagcatccc aaagcccccct ctgctcaacc cagcccagtt 3223

ttgca
3228

<210> 3
<211> 1375
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (490)..(1044)

<400> 3
tgcttccgcg tctagccgga gaaacttgag ccggctgccc cgcccacggt gcccgaagcc 60

ccaaaggctg gaattagggg ctagaagtct ggcacccacc gcctggccag gtgttcggga 120

cgcgaccagg tgggcggtcg ccccgccccg ggagcgcggc ttaatagctg agagcccggg 180

ggccaggccg cggctgcggc ccaggcaacg ccctgagggt ggccacgctg ccaggtgttc 240

cactcccccg ggactatggg caagggcccg gggcggggag ggcggcaggt gctgacactg 300

gagctgcgcc ggaggtcggg gaactcggcc tcctaagact gaggacactc gcctgctggg 360

ccggtcgagc tgtgcggtgc cctccgggac gcagggggcg ctgcagccac gctgggtcag 420

gctccgcagg gccctcccaa cccggggact aacggcgccg gtgacgactt cgccgcgcgt 480

tggtcagcc atg gcc acc gct ctc gcg cta cgt agc ttg tac cga gcg 528
Met Ala Thr Ala Leu Ala Leu Arg Ser Leu Tyr Arg Ala
1 5 10

cga ccc tcg ctg cgc tgt ccg ccc gtt gag ctt ccc tgg gcc ccg cgg 576
Arg Pro Ser Leu Arg Cys Pro Pro Val Glu Leu Pro Trp Ala Pro Arg
15 20 25

cga ggg cat cgg ctc tcg ccg gcg gat gac gag ctg tat cag cgg acg 624
Arg Gly His Arg Leu Ser Pro Ala Asp Asp Glu Leu Tyr Gln Arg Thr
30 35 40 45

cgc atc tct ctg ctg caa cgc gag gcc gct cag gca atg tac atc gac 672
Arg Ile Ser Leu Leu Gln Arg Glu Ala Ala Gln Ala Met Tyr Ile Asp
50 55 60

agc tac aac agc cgc ggc ttc atg ata aac gga aac cgc gtg ctc ggc 720
Ser Tyr Asn Ser Arg Gly Phe Met Ile Asn Gly Asn Arg Val Leu Gly
65 70 75

ccc tgc gct ctg ctc ccg cac tcg gtg gtg cag tgg aac gtg gga tcc 768
Pro Cys Ala Leu Leu Pro His Ser Val Val Gln Trp Asn Val Gly Ser
80 85 90

cac cag gac atc acc gaa gac agc ttt tcc ctc ttc tgg ttg ctg gag 816

His Gln Asp Ile Thr Glu Asp Ser Phe Ser Leu Phe Trp Leu Leu Glu
95 100 105

ccc cgg ata gag atc gtg gtg gtg ggg act gga gac cgg acc gag agg 864
Pro Arg Ile Glu Ile Val Val Val Gly Thr Gly Asp Arg Thr Glu Arg
110 115 120 125

ctg cag tcc cag gtg ctt caa gcc atg agg cag cgg ggc att gct gtg 912
Leu Gln Ser Gln Val Leu Gln Ala Met Arg Gln Arg Gly Ile Ala Val
130 135 140

gaa gtg cag gac acg ccc aat gcc tgt gcc acc ttc aac ttc ctg tgt 960
Glu Val Gln Asp Thr Pro Asn Ala Cys Ala Thr Phe Asn Phe Leu Cys
145 150 155

cat gaa ggc cga gta act gga gct gct ctc atc cct cca cca gga ggg 1008
His Glu Gly Arg Val Thr Gly Ala Ala Leu Ile Pro Pro Pro Gly Gly
160 165 170

act tca ctt aca tct ttg ggc caa gct gct caa tga accg ccaggaactg 1058
Thr Ser Leu Thr Ser Leu Gly Gln Ala Ala Gln
175 180

acctgctgac tgcactctgc caggcttccc aatgctttca ctcttatcta ccctttggca 1118

cttatcttgc ttatcaacat aataatttat acacttctcc cattttgtat caggtgtgtt 1178

gctggccagg agctgatggc tcactgggct cttggagggg aatgtgaaga aaccaaggag 1238

tcactttttc atctagatta cttaggattc cttgactttt cagaagtcgg gaagcagtat 1298

gtttgcctgt tgtagaccta cttgctcaca tgcagatttg agaggacctc aacggctttt 1358

ctcacaaaaa aaaaaaa 1375

<210> 4
<211> 569
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (194)..(562)

<400> 4
aacacctgct ttttggggtg gttgtgagga ggcaaacagt gtccagcagg cctgccccag 60

aaccccactt catccaaggg ggctgggagg tctctgatga acccaaccat gtgccttcac 120

agccccgtca tgtgcgtctc caggtgtgag acaatgggat gtaatgacca ttgttctcga 180

tgccagtaaa gcc atg cct ggg cac atg cca aga gga ggt tgg tgt ctg 229
Met Pro Gly His Met Pro Arg Gly Gly Trp Cys Leu
1 5 10

ttt aca gga cct gct tgg gtt ggc agt tcc cat ggg gaa ggg aag agg 277
Phe Thr Gly Pro Ala Trp Val Gly Ser Ser His Gly Glu Gly Lys Arg
15 20 25

cgt ggt ggg agc ttg gcc cag ctt ggc aaa caa agc cag tgg gaa cac 325
Arg Gly Gly Ser Leu Ala Gln Leu Gly Lys Gln Ser Gln Trp Glu His
30 35 40

aag cct gtc ctc ctc cat ggc cag ggg tgg aaa aaa gag tgt gta aag 373
Lys Pro Val Leu Leu His Gly Gln Gly Trp Lys Lys Glu Cys Val Lys
45 50 55 60

cca agc cca atg ccc ctc ttt cca tct gtc ccc cag atg tcc tac gct 421
Pro Ser Pro Met Pro Leu Phe Pro Ser Val Pro Gln Met Ser Tyr Ala
65 70 75

tcg aca aca cct ata gct ttg tcc acg cca aga agg tca gct tca cag 469
Ser Thr Thr Pro Ile Ala Leu Ser Thr Pro Arg Arg Ser Ala Ser Gln
80 85 90

tgg agg tcc tgc tcc ctg acg agg gca tgc aga aat atg ata agg agc 517
Trp Arg Ser Cys Ser Leu Thr Arg Ala Cys Arg Asn Met Ile Arg Ser
95 100 105

tca ccc ctg tct agg tgg ctc cct cat ttc tca gag acc tcc taa ccc 565
Ser Pro Leu Ser Arg Trp Leu Pro His Phe Ser Glu Thr Ser
110 115 120

tttg 569

<210> 5
<211> 1881
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1584)

<400> 5
atg gcc cag aca ctg cag atg gag atc ccg aac ttc ggc aac agc atc 48
Met Ala Gln Thr Leu Gln Met Glu Ile Pro Asn Phe Gly Asn Ser Ile
1 5 10 15

ctg gag tgc ctc aat gaa cag cgg ctg cag ggc ctg tac tgt gac gtg 96
Leu Glu Cys Leu Asn Glu Gln Arg Leu Gln Gly Leu Tyr Cys Asp Val
20 25 30

tca gtg gtg gtc aag ggc cat gcc ttc aag gcc cac cgg gcc gtg ctt 144
Ser Val Val Val Lys Gly His Ala Phe Lys Ala His Arg Ala Val Leu
35 40 45

```
 gct gcc agc agc tcc tac ttc cgg gac ctg ttc aac aac agc cgc
agc      192
 Ala Ala Ser Ser Ser Tyr Phe Arg Asp Leu Phe Asn Asn Ser Arg
Ser
      50                  55                  60

 gcc gtg gtg gag ctg ccg gcg gct gtg cag ccc cag tct ttc cag
cag      240
 Ala Val Val Glu Leu Pro Ala Ala Val Gln Pro Gln Ser Phe Gln
Gln
  65                  70                  75
80

 atc ctc agc ttc tgc tac acg ggc cgg ctg agc atg aac gtg ggc
gac      288
 Ile Leu Ser Phe Cys Tyr Thr Gly Arg Leu Ser Met Asn Val Gly
Asp
                 85                  90                  95

 cag ttc ctg ctc atg tac acg gct ggc ttc ctg cag atc cag gag
atc      336
 Gln Phe Leu Leu Met Tyr Thr Ala Gly Phe Leu Gln Ile Gln Glu
Ile
             100                 105                 110

 atg gag aag ggc acc gag ttc ttc ctc aag gtg agc tcc ccg agc
tgc      384
 Met Glu Lys Gly Thr Glu Phe Phe Leu Lys Val Ser Ser Pro Ser
Cys
         115                 120                 125

 gac tcc cag ggc ctg cat gcg gag gag gcc cca tcg tcg gag ccc
cag      432
 Asp Ser Gln Gly Leu His Ala Glu Glu Ala Pro Ser Ser Glu Pro
Gln
     130                 135                 140

 agc ccc gtg gcg cag aca tcg ggc tgg cca gcc tgt agc acc ccg
ctg      480
 Ser Pro Val Ala Gln Thr Ser Gly Trp Pro Ala Cys Ser Thr Pro
Leu
 145                 150                 155
160

 ccc ctc gtg tcg cgg gtg aag acg gag cag cag gag tcg gac tcc
gtg      528
 Pro Leu Val Ser Arg Val Lys Thr Glu Gln Gln Glu Ser Asp Ser
Val
                 165                 170                 175

 cag tgc atg ccc gtg gcc aag cgg ctg tgg gac agt ggc cag aag
gag      576
```

Gln Cys Met Pro Val Ala Lys Arg Leu Trp Asp Ser Gly Gln Lys Glu
180 185 190

gct ggg ggc ggc ggc aat ggc agc cgc aag atg gcc aag ttc tcc acg 624
Ala Gly Gly Gly Gly Asn Gly Ser Arg Lys Met Ala Lys Phe Ser Thr
195 200 205

ccg gac ctg gct gcc aac cgg cct cac cag ccc ccg cca ccc caa cag 672
Pro Asp Leu Ala Ala Asn Arg Pro His Gln Pro Pro Pro Pro Gln Gln
210 215 220

gct ccg gtg gtg gca gca gcc cag ccc gcc gtg gct gcg gga gca ggg 720
Ala Pro Val Val Ala Ala Ala Gln Pro Ala Val Ala Ala Gly Ala Gly
225 230 235 240

cag cca gcc ggt ggg gtg gca gca gca ggg ggt gtg gtg agt ggg ccc 768
Gln Pro Ala Gly Gly Val Ala Ala Ala Gly Gly Val Val Ser Gly Pro
245 250 255

agc acg tcg gag cgg acc agc cca ggc acc tca agc gcc tac acc agc 816
Ser Thr Ser Glu Arg Thr Ser Pro Gly Thr Ser Ser Ala Tyr Thr Ser
260 265 270

gac agc cct ggc tcc tac cac aat gag gag gac gag gag gag gat ggt 864
Asp Ser Pro Gly Ser Tyr His Asn Glu Glu Asp Glu Glu Glu Asp Gly
275 280 285

ggc gag gag ggc atg gat gag cag tac cgg cag atc tgc aac atg tac 912
Gly Glu Glu Gly Met Asp Glu Gln Tyr Arg Gln Ile Cys Asn Met Tyr
290 295 300

acc atg tac agc atg atg aac gtc ggc cag aca gcc gag aag gtg gag 960
Thr Met Tyr Ser Met Met Asn Val Gly Gln Thr Ala Glu Lys Val Glu

305 310 315
320

gcc ctc ccg gag cag gta gcc ccc gag tcc cga aat cgc atc cgg
gtt 1008
Ala Leu Pro Glu Gln Val Ala Pro Glu Ser Arg Asn Arg Ile Arg
Val
325 330 335

cgg caa gac ctg gcg tct ctc ccg gct gaa ctt atc aac cag att
ggg 1056
Arg Gln Asp Leu Ala Ser Leu Pro Ala Glu Leu Ile Asn Gln Ile
Gly
340 345 350

aac cgc tgc cac ccc aag ctc tac gac gag ggc gac ccc tct gag
aag 1104
Asn Arg Cys His Pro Lys Leu Tyr Asp Glu Gly Asp Pro Ser Glu
Lys
355 360 365

ctg gag ctg gtg aca ggc acc aac gtg tac atc aca agg gcg cag
ctg 1152
Leu Glu Leu Val Thr Gly Thr Asn Val Tyr Ile Thr Arg Ala Gln
Leu
370 375 380

atg aac tgc cac gtc agc gca ggc acg cgg cac aag gtc cta ctg
cgg 1200
Met Asn Cys His Val Ser Ala Gly Thr Arg His Lys Val Leu Leu
Arg
385 390 395
400

cgg ctc ctg gcc tcc ttc ttt gac cgg aac acg ctg gcc aac agc
tgc 1248
Arg Leu Leu Ala Ser Phe Phe Asp Arg Asn Thr Leu Ala Asn Ser
Cys
405 410 415

ggc acc ggc atc cgc tct tct acc aac gat ccc cgt cgg aag ccc
ctg 1296
Gly Thr Gly Ile Arg Ser Ser Thr Asn Asp Pro Arg Arg Lys Pro
Leu
420 425 430

gac agc cgc gtg ctc cac gct gtc aag tac tac tgc cag aac ttc
gcc 1344
Asp Ser Arg Val Leu His Ala Val Lys Tyr Tyr Cys Gln Asn Phe
Ala
435 440 445

```
 ccc aac ttc aag gag agc gag atg aat gcc atc gcg gcc gac atg
tgc     1392
 Pro Asn Phe Lys Glu Ser Glu Met Asn Ala Ile Ala Ala Asp Met
Cys
    450                 455                 460

 acc aac gcc cgc cgc gtc gtg cgc aag agc tgg atg ccc aag gtc
aag     1440
 Thr Asn Ala Arg Arg Val Val Arg Lys Ser Trp Met Pro Lys Val
Lys
 465                470                 475
480

 gtg ctc aag gct gag gat gac gcc tac acc acc ttc atc agt gaa
acg     1488
 Val Leu Lys Ala Glu Asp Asp Ala Tyr Thr Thr Phe Ile Ser Glu
Thr
                485                 490                 495

 ggc aag atc gag ccg gac atg atg ggt gtg gag cat ggc ttc gag
acc     1536
 Gly Lys Ile Glu Pro Asp Met Met Gly Val Glu His Gly Phe Glu
Thr
            500                 505                 510

 gcc agc cac gag ggc gag gcg ggt ccc tcg gct gaa gcc ctg cag
taa     1584
 Ala Ser His Glu Gly Glu Ala Gly Pro Ser Ala Glu Ala Leu Gln
        515                 520                 525

 cccgcccagc ctcccgcggg gccgcacact tcccctccca acacacacac
acacctgcca   1644

 tcttggtcat gagctactgt ctgtccctcc ccaggacccg cggtgggtgc
tgcatgttcc   1704

 cggccctctg cccctcctgt cctacccccct ttccccaccg agagctgggc
cgggagagga   1764

 ccgcagggca ggtggcgtga ggtccgtgtt gccttcttta acacacactc
gtgcagtggg   1824

 ggagttctgg ctccccaacc taacccctag ccgtcatctc cacctaaaaa aaaaaaa
1881



      <210> 6
      <211> 2889
      <212> DNA
      <213> Homo sapiens
```

<220>
<221> CDS
<222> (14)..(1744)

<400> 6
ctgtcctgga aag atg cta gca atg ggg gcg ctg gca gga ttc tgg atc 49
Met Leu Ala Met Gly Ala Leu Ala Gly Phe Trp Ile
1 5 10

ctc tgc ctc ctc act tat ggt tac ctg tcc tgg ggc cag gcc tta gaa 97
Leu Cys Leu Leu Thr Tyr Gly Tyr Leu Ser Trp Gly Gln Ala Leu Glu
15 20 25

gag gag gaa gaa ggg gcc tta cta gct caa gct gga gag aaa cta gag 145
Glu Glu Glu Glu Gly Ala Leu Leu Ala Gln Ala Gly Glu Lys Leu Glu
30 35 40

ccc agc aca act tcc acc tcc cag ccc cat ctc att ttc atc cta gcg 193
Pro Ser Thr Thr Ser Thr Ser Gln Pro His Leu Ile Phe Ile Leu Ala
45 50 55 60

gat gat cag gga ttt aga gat gtg ggt tac cac gga tct gag att aaa 241
Asp Asp Gln Gly Phe Arg Asp Val Gly Tyr His Gly Ser Glu Ile Lys
65 70 75

aca cct act ctt gac aag ctc gct gcc gaa gga gtt aaa ctg gag aac 289
Thr Pro Thr Leu Asp Lys Leu Ala Ala Glu Gly Val Lys Leu Glu Asn
80 85 90

tac tat gtc cag cct att tgc aca cca tcc agg agt cag ttt att act 337
Tyr Tyr Val Gln Pro Ile Cys Thr Pro Ser Arg Ser Gln Phe Ile Thr
95 100 105

gga aag tat cag ata cac acc gga ctt caa cat tct atc ata aga cct 385

Gly Lys Tyr Gln Ile His Thr Gly Leu Gln His Ser Ile Ile Arg
Pro
110 115 120

acc caa ccc aac tgt tta cct ctg gac aat gcc acc cta cct cag
aaa 433
Thr Gln Pro Asn Cys Leu Pro Leu Asp Asn Ala Thr Leu Pro Gln
Lys
125 130 135
140

ctg aag gag gtt gga tat tca acg cat atg gtc gga aaa tgg cac
ttg 481
Leu Lys Glu Val Gly Tyr Ser Thr His Met Val Gly Lys Trp His
Leu
145 150 155

ggt ttt tac aga aaa gaa tgc atg ccc acc aga aga gga ttt gat
acc 529
Gly Phe Tyr Arg Lys Glu Cys Met Pro Thr Arg Arg Gly Phe Asp
Thr
160 165 170

ttt ttt ggt tcc ctt ttg gga agt ggg gat tac tat aca cac tac
aaa 577
Phe Phe Gly Ser Leu Leu Gly Ser Gly Asp Tyr Tyr Thr His Tyr
Lys
175 180 185

tgt gac agt cct ggg atg tgt ggc tat gac ttg tat gaa aac gac
aat 625
Cys Asp Ser Pro Gly Met Cys Gly Tyr Asp Leu Tyr Glu Asn Asp
Asn
190 195 200

gct gcc tgg gac tat gac aat ggc ata tac tcc aca cag atg tac
act 673
Ala Ala Trp Asp Tyr Asp Asn Gly Ile Tyr Ser Thr Gln Met Tyr
Thr
205 210 215
220

cag aga gta cag caa atc tta gct tcc cat aac ccc aca aag cct
ata 721
Gln Arg Val Gln Gln Ile Leu Ala Ser His Asn Pro Thr Lys Pro
Ile
225 230 235

ttt tta tat att gcc tat caa gct gtt cat tca cca ctg caa gct
cct 769
Phe Leu Tyr Ile Ala Tyr Gln Ala Val His Ser Pro Leu Gln Ala
Pro

240 245 250

ggc agg tat ttc gaa cac tac cga tcc att atc aac ata aac agg agg 817
Gly Arg Tyr Phe Glu His Tyr Arg Ser Ile Ile Asn Ile Asn Arg Arg
255 260 265

aga tat gct gcc atg ctt tcc tgc tta gat gaa gca atc aac aac gtg 865
Arg Tyr Ala Ala Met Leu Ser Cys Leu Asp Glu Ala Ile Asn Asn Val
270 275 280

aca ttg gct cta aag act tat ggt ttc tat aac aac agc att atc att 913
Thr Leu Ala Leu Lys Thr Tyr Gly Phe Tyr Asn Asn Ser Ile Ile Ile
285 290 295 300

tac tct tca gat aat ggt ggc cag cct acg gca gga ggg agt aac tgg 961
Tyr Ser Ser Asp Asn Gly Gly Gln Pro Thr Ala Gly Gly Ser Asn Trp
305 310 315

cct ctc aga ggt agc aaa gga aca tat tgg gaa gga ggg atc cgg gct 1009
Pro Leu Arg Gly Ser Lys Gly Thr Tyr Trp Glu Gly Gly Ile Arg Ala
320 325 330

gta ggc ttt gtg cat agc cca ctt ctg aaa aac aag gga aca gtg tgt 1057
Val Gly Phe Val His Ser Pro Leu Leu Lys Asn Lys Gly Thr Val Cys
335 340 345

aag gaa ctt gtg cac atc act gac tgg tac ccc act ctc att tca ctg 1105
Lys Glu Leu Val His Ile Thr Asp Trp Tyr Pro Thr Leu Ile Ser Leu
350 355 360

gct gaa gga cag att gat gag gac att caa cta gat ggc tat gat atc 1153
Ala Glu Gly Gln Ile Asp Glu Asp Ile Gln Leu Asp Gly Tyr Asp Ile
365 370 375 380

tgg gag acc ata agt gag ggt ctt cgc tca ccc cga gta gat att ttg 1201
Trp Glu Thr Ile Ser Glu Gly Leu Arg Ser Pro Arg Val Asp Ile Leu
385 390 395

cat aac att gac ccc ata tac acc aag gca aaa aat ggc tcc tgg gca 1249
His Asn Ile Asp Pro Ile Tyr Thr Lys Ala Lys Asn Gly Ser Trp Ala
400 405 410

gca ggc tat ggg atc tgg aac act gca atc cag tca gcc atc aga gtg 1297
Ala Gly Tyr Gly Ile Trp Asn Thr Ala Ile Gln Ser Ala Ile Arg Val
415 420 425

cag cac tgg aaa ttg ctt aca gga aat cct ggc tac agc gac tgg gtc 1345
Gln His Trp Lys Leu Leu Thr Gly Asn Pro Gly Tyr Ser Asp Trp Val
430 435 440

ccc cct cag tct ttc agc aac ctg gga ccg aac cgg tgg cac aat gaa 1393
Pro Pro Gln Ser Phe Ser Asn Leu Gly Pro Asn Arg Trp His Asn Glu
445 450 455
460

cgg atc acc ttg tca act ggc aaa agt gta tgg ctt ttc aac atc aca 1441
Arg Ile Thr Leu Ser Thr Gly Lys Ser Val Trp Leu Phe Asn Ile Thr
465 470 475

gcc gac cca tat gag agg gtg gac cta tct aac agg tat cca gga atc 1489
Ala Asp Pro Tyr Glu Arg Val Asp Leu Ser Asn Arg Tyr Pro Gly Ile
480 485 490

gtg aag aag ctc cta cgg agg ctc tca cag ttc aac aaa act gca gtg 1537
Val Lys Lys Leu Leu Arg Arg Leu Ser Gln Phe Asn Lys Thr Ala Val
495 500 505

ccg gtc agg tat ccc ccc aaa gac ccc aga agt aac cct agg ctc aat 1585

Pro Val Arg Tyr Pro Pro Lys Asp Pro Arg Ser Asn Pro Arg Leu
Asn
510 515 520

gga ggg gtc tgg gga cca tgg tat aaa gag gaa acc aag aaa aag
aag 1633
Gly Gly Val Trp Gly Pro Trp Tyr Lys Glu Glu Thr Lys Lys Lys
Lys
525 530 535
540

cca agc aaa aat cag gct gag aaa aag caa aag aaa agc aaa aaa
aag 1681
Pro Ser Lys Asn Gln Ala Glu Lys Lys Gln Lys Lys Ser Lys Lys
Lys
545 550 555

aag aag aaa cag cag aaa gca gtc tca ggt tca act tgc cat tca
ggt 1729
Lys Lys Lys Gln Gln Lys Ala Val Ser Gly Ser Thr Cys His Ser
Gly
560 565 570

gtt act tgt gga taa gcacaaatat ttcctgtttg gttaaacttt
aatcagttct 1784
Val Thr Cys Gly
575

tatctttcat ctgtttccta ggtaaaccag caaatttggc tcgataatat
cgctggccta 1844

agcgtcaggc ttgttttcat gctgtgccac tccagagact tctgccacct
ggccgccaca 1904

ctgaaaactg tcctgctcag tgccaaggtg ctactcttgc aagccacact
tagagagagt 1964

ggagatgttt atttctctcg ctcctttaga aaacgtggtg agtcctgagt
tccactgctg 2024

tgcttcagtc aactgaccaa acactgcttt gaattatagg aggagaacaa
taacctacca 2084

tccgcaagca tgctaatttg atggaagtta cagggtagca tgattaaaac
tacctttgat 2144

aaattacagt caaagattgt gtcacctcaa aggccttgaa gaatatattt
tcttggtgaa 2204

tttttgtatg tctgtcatat gacacttggg ttttttaatt aattctattt
tatatatata 2264

aatatatgtt tcttttcctg tgaaaagctg tttttctcac atgtgaacag cttgcacctc 2324

attttaccat gcgtgaggga atggcaaata agaatgtttg agcacactgc ccacaatgaa 2384

tgtaactatt ttctaaacac tttactagaa gaacatttca gtataaaaaa cctaatttat 2444

ttttacagaa aaatattttg ttgtttttat aaaaagttat gcaaatgact tttattttta 2504

tttcctgcat accattagaa gaattttatt tcatttcttc aaattatcaa gcactgtaat 2564

acctataaat taatgtaata ctgtgtgaat tcagactatt aaaaacatca ttcagaaaac 2624

tttataatcg tcattgttca atcaagattt tgaatgtaat aagatgaata tattccttac 2684

aaattacttg gaaattcaat gtttgtgcag agttgagaca actttattgt ttctatcata 2744

aactatttat gtatcttaat tattaaaatg atttacttta tggcactaga aaatttactg 2804

tggcttttct gatctaactt ctagctaaaa ttgtatcatt ggccctaaaa aataaaaatc 2864

tttactaata ggcaaaaaaa aaaaa 2889

<210> 7
<211> 1356
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (51)..(1061)

<400> 7
cggacgcgtg ggccgggggt tcgcccgcgg aggccgggga gcagccgacc atg gag 56
Met Glu
1

ccc cag aaa atc atg cca ccc tca aag cct cat cca cct gtc gtg ggc 104
Pro Gln Lys Ile Met Pro Pro Ser Lys Pro His Pro Pro Val Val Gly
5 10 15

aaa gtg act cat cac agc att gaa tta tac tgg gat ctg gaa aag aaa 152
Lys Val Thr His His Ser Ile Glu Leu Tyr Trp Asp Leu Glu Lys Lys
20 25 30

gcc aaa cgc caa gga cct caa gag cag tgg ttc agg ttc tcg att gaa 200
Ala Lys Arg Gln Gly Pro Gln Glu Gln Trp Phe Arg Phe Ser Ile Glu
35 40 45 50

gaa gaa gac ccc aaa atg cac act tat ggt atc att tat acg gga tat 248
Glu Glu Asp Pro Lys Met His Thr Tyr Gly Ile Ile Tyr Thr Gly Tyr
55 60 65

gca acg aag cat gtt gtt gaa ggt ctg gaa cca agg acg ctg tac aga 296
Ala Thr Lys His Val Val Glu Gly Leu Glu Pro Arg Thr Leu Tyr Arg
70 75 80

ttt cgc ctg aag gtc acc agc ccc tct ggg gag tgt gag tac agc cca 344
Phe Arg Leu Lys Val Thr Ser Pro Ser Gly Glu Cys Glu Tyr Ser Pro
85 90 95

ctc gtc tca gtg tct aca acc aga gag ccc ata agt agt gag cac ttg 392
Leu Val Ser Val Ser Thr Thr Arg Glu Pro Ile Ser Ser Glu His Leu
100 105 110

cac cgg gct gtc agt gtg aat gat gaa gat ttg ctg gtc cga ata ctt 440
His Arg Ala Val Ser Val Asn Asp Glu Asp Leu Leu Val Arg Ile Leu
115 120 125 130

caa gga ggc cgt gtt aag gtt gat gtt ccc aat aag ttt ggc ttt acc 488

Gln Gly Gly Arg Val Lys Val Asp Val Pro Asn Lys Phe Gly Phe Thr
135 140 145

gct ctg atg gtt gct gcc cag aaa gga tac acc agg ctt gtg aaa atc 536
Ala Leu Met Val Ala Ala Gln Lys Gly Tyr Thr Arg Leu Val Lys Ile
150 155 160

cta gtt tct aat ggc aca gac gtg aat ctg aag aat gga agt ggc aag 584
Leu Val Ser Asn Gly Thr Asp Val Asn Leu Lys Asn Gly Ser Gly Lys
165 170 175

gac agt cta atg ctg gcg tgc tat gcg gga cac cta gat gtt gtg aaa 632
Asp Ser Leu Met Leu Ala Cys Tyr Ala Gly His Leu Asp Val Val Lys
180 185 190

tat ctc cga aga cat ggc gct tct tgg cag gct aga gac ctg gga ggc 680
Tyr Leu Arg Arg His Gly Ala Ser Trp Gln Ala Arg Asp Leu Gly Gly
195 200 205 210

tgt aca gct ctg cac tgg gct gca gat gga ggc cac tgc agt gtg att 728
Cys Thr Ala Leu His Trp Ala Ala Asp Gly Gly His Cys Ser Val Ile
215 220 225

gag tgg atg ata aag gat ggc tgt gag gta gac gtc gtg gac act ggt 776
Glu Trp Met Ile Lys Asp Gly Cys Glu Val Asp Val Val Asp Thr Gly
230 235 240

tca gga tgg acc cca ctc atg aga gtc tct gcg gtg tcg gga aat cag 824
Ser Gly Trp Thr Pro Leu Met Arg Val Ser Ala Val Ser Gly Asn Gln
245 250 255

agg gtg gcc tct ctt cta att gat gct ggg gcc aat gtg aat gtg aag 872
Arg Val Ala Ser Leu Leu Ile Asp Ala Gly Ala Asn Val Asn Val Lys
260 265 270

gac aga aat gga aag acg ccc ctt atg gtg gct gtg tta aat aat cat 920
Asp Arg Asn Gly Lys Thr Pro Leu Met Val Ala Val Leu Asn Asn His
275 280 285 290

gaa gag tta gtt cag tta ctt ctt gac aaa ggg gca gat gca agt gta 968
Glu Glu Leu Val Gln Leu Leu Leu Asp Lys Gly Ala Asp Ala Ser Val
295 300 305

aaa aat gag ttc ggc aaa ggt gtc cta gaa atg gcc aga gtt ttt gac 1016
Lys Asn Glu Phe Gly Lys Gly Val Leu Glu Met Ala Arg Val Phe Asp
310 315 320

aga cag gtt ggg atg ctc ttt ctg cct atc aag gct aat ttt tag tcc 1064
Arg Gln Val Gly Met Leu Phe Leu Pro Ile Lys Ala Asn Phe
325 330 335

ttctagactc agaaaaacca ccaaaaccct ggattacatt cgcctgtctg ttttgtttat 1124

ctagagtgta gtctccttat tagaagaaag gaaaaaaaag cagaggccaa agaagtcttg 1184

tgtctgctga tgagagcacc actcatctgc gaaacgcacg taaaacaaag tgaaccgtga 1244

ctgttaaact agggatggga aattctgcat cttggggggc tgtacattta tttatttagt 1304

tgaagattca ctgatcccac tttgaaatac atctttttac ctaaaaaaaa aa 1356

<210> 8
<211> 2124
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1866)

<400> 8
atg gcg ctg ggc ttg gag cag gcg gag gag cag cgg ttg tac cag cag 48
Met Ala Leu Gly Leu Glu Gln Ala Glu Glu Gln Arg Leu Tyr Gln Gln
1 5 10 15

acg ctc ctg caa gac ggg ctc aaa gac atg ctg gac cat ggc aag ttc 96
Thr Leu Leu Gln Asp Gly Leu Lys Asp Met Leu Asp His Gly Lys Phe
20 25 30

ctc gac tgt gtg gtg cgg gcg ggc gag cgc gag ttc ccg tgc cat cgc 144
Leu Asp Cys Val Val Arg Ala Gly Glu Arg Glu Phe Pro Cys His Arg
35 40 45

ctg gtg ctg gcc gcc tgc agc ccc tac ttc cgg gcg cgc ttt cta gcc 192
Leu Val Leu Ala Ala Cys Ser Pro Tyr Phe Arg Ala Arg Phe Leu Ala
50 55 60

gag ccg gag cgc gcg ggc gag ctg cac ctg gag gag gtg tcc ccg gac 240
Glu Pro Glu Arg Ala Gly Glu Leu His Leu Glu Glu Val Ser Pro Asp
65 70 75 80

gtg gtg gcc cag gtg ctg cac tac ctg tac aca tca gag atc gcg ctg 288
Val Val Ala Gln Val Leu His Tyr Leu Tyr Thr Ser Glu Ile Ala Leu
85 90 95

gat gag gcg agc gtg cag gat ttg ttc gcc gcg gca cac cgc ttc cag 336
Asp Glu Ala Ser Val Gln Asp Leu Phe Ala Ala Ala His Arg Phe Gln
100 105 110

atc cct tcc atc ttc acc atc tgc gtg tcc ttc ctg cag aag cgc ctg 384
Ile Pro Ser Ile Phe Thr Ile Cys Val Ser Phe Leu Gln Lys Arg Leu
115 120 125

tgc ctc tcc aac tgc ttg gcc gtc ttc cgt ctc ggc ctc ctg ctc gac 432

Cys Leu Ser Asn Cys Leu Ala Val Phe Arg Leu Gly Leu Leu Leu
Asp
130 135 140

tgc gcg cgt ctc gcc gtg gct gcc cgc gac ttc atc tgc gct cac
ttc 480
Cys Ala Arg Leu Ala Val Ala Ala Arg Asp Phe Ile Cys Ala His
Phe
145 150 155
160

acg ctg gtg gcg cgc gac gct gac ttc ctc gga ctc tcg gcc gac
gag 528
Thr Leu Val Ala Arg Asp Ala Asp Phe Leu Gly Leu Ser Ala Asp
Glu
165 170 175

ctc atc gcc atc atc tcc agc gac ggc ctt aac gtg gag aag gag
gag 576
Leu Ile Ala Ile Ile Ser Ser Asp Gly Leu Asn Val Glu Lys Glu
Glu
180 185 190

gca gtg ttc gag gcg gtg atg cgg tgg gcg ggt agc ggc gac gcc
gag 624
Ala Val Phe Glu Ala Val Met Arg Trp Ala Gly Ser Gly Asp Ala
Glu
195 200 205

gcg cag gct gag cgc cag cgc gcg ctg ccc acc gtc ttc gag agc
gtg 672
Ala Gln Ala Glu Arg Gln Arg Ala Leu Pro Thr Val Phe Glu Ser
Val
210 215 220

cgc tgc cgc ttg ctg ccg cgc gcc ttt ctg gaa agc cgc gtg gag
cgc 720
Arg Cys Arg Leu Leu Pro Arg Ala Phe Leu Glu Ser Arg Val Glu
Arg
225 230 235
240

cac cct ctc gtg cgt gcc cag ccc gag ttg ctg cgc aag gtg cag
atg 768
His Pro Leu Val Arg Ala Gln Pro Glu Leu Leu Arg Lys Val Gln
Met
245 250 255

gtg aag gat gca cac gag ggc cgc atc acc acg ctg cgg aag aaa
aag 816
Val Lys Asp Ala His Glu Gly Arg Ile Thr Thr Leu Arg Lys Lys
Lys

260 265 270

aag ggg aag gat gga gcc ggg gcc aag gag gct gat aag ggc aca agc 864
Lys Gly Lys Asp Gly Ala Gly Ala Lys Glu Ala Asp Lys Gly Thr Ser
275 280 285

aaa gcc aaa gca gag gag gat gag gag gcc gaa cgt atc ctt cct ggg 912
Lys Ala Lys Ala Glu Glu Asp Glu Glu Ala Glu Arg Ile Leu Pro Gly
290 295 300

atc ctc aat gac acc ctg cgc ttc ggc atg ttc ctg cag gat ctc atc 960
Ile Leu Asn Asp Thr Leu Arg Phe Gly Met Phe Leu Gln Asp Leu Ile
305 310 315 320

ttc atg atc agt gag gag ggc gct gtg gcc tac gat cca gca gcc aac 1008
Phe Met Ile Ser Glu Glu Gly Ala Val Ala Tyr Asp Pro Ala Ala Asn
325 330 335

gag tgc tac tgt gct tcc ctc tcc aac cag gtc ccc aag aac cac gtc 1056
Glu Cys Tyr Cys Ala Ser Leu Ser Asn Gln Val Pro Lys Asn His Val
340 345 350

agc ctg gtt acc aag gag aac cag gtc ttc gtg gct gga ggc ctc ttc 1104
Ser Leu Val Thr Lys Glu Asn Gln Val Phe Val Ala Gly Gly Leu Phe
355 360 365

tac aac gaa gac aac aaa gag gac ccc atg agc gca tac ttc ctg cag 1152
Tyr Asn Glu Asp Asn Lys Glu Asp Pro Met Ser Ala Tyr Phe Leu Gln
370 375 380

ttt gac cat ctg gac tca gag tgg ctg ggg atg cca ccg ctg ccc tcg 1200
Phe Asp His Leu Asp Ser Glu Trp Leu Gly Met Pro Pro Leu Pro Ser
385 390 395 400

ccc cgc tgc ctc ttt ggc ctg gga gaa gct ctc aac tcc atc tac gtg 1248
Pro Arg Cys Leu Phe Gly Leu Gly Glu Ala Leu Asn Ser Ile Tyr Val
405 410 415

gtc ggt ggc aga gag atc aag gac ggc gag cgc tgc ctg gac tcg gtc 1296
Val Gly Gly Arg Glu Ile Lys Asp Gly Glu Arg Cys Leu Asp Ser Val
420 425 430

atg tgc tac gac agg ctg tca ttc aaa tgg ggt gaa tcg gac ccg ctg 1344
Met Cys Tyr Asp Arg Leu Ser Phe Lys Trp Gly Glu Ser Asp Pro Leu
435 440 445

cct tac gtg gtg tat ggc cac aca gtg ctc tcc cac atg gac ctt gtc 1392
Pro Tyr Val Val Tyr Gly His Thr Val Leu Ser His Met Asp Leu Val
450 455 460

tac gta att ggc ggc aaa ggc agt gac agg aag tgc ctg aac aag atg 1440
Tyr Val Ile Gly Gly Lys Gly Ser Asp Arg Lys Cys Leu Asn Lys Met
465 470 475 480

tgc gtc tat gac ccc aag aag ttt gag tgg aag gag ctg gca ccc atg 1488
Cys Val Tyr Asp Pro Lys Lys Phe Glu Trp Lys Glu Leu Ala Pro Met
485 490 495

cag acc gcc cgc tca ctc ttt ggg gcc act gtc cat gat ggc cgc att 1536
Gln Thr Ala Arg Ser Leu Phe Gly Ala Thr Val His Asp Gly Arg Ile
500 505 510

atc gtg gca gct ggg gtc acc gac aca ggg ctg acc agt tct gcc gaa 1584
Ile Val Ala Ala Gly Val Thr Asp Thr Gly Leu Thr Ser Ser Ala Glu
515 520 525

gtg tac agc atc aca gac aac aag tgg gca ccc ttc gag gcc ttc cca 1632

Val Tyr Ser Ile Thr Asp Asn Lys Trp Ala Pro Phe Glu Ala Phe Pro
530 535 540

cag gag cgt agc tca ctc agc ctg gtc agc ctg gtg ggt acc ctc tat 1680
Gln Glu Arg Ser Ser Leu Ser Leu Val Ser Leu Val Gly Thr Leu Tyr
545 550 555 560

gcc att ggt ggc ttt gcc aca ctg gag acg gag tct gga gag ctg gtt 1728
Ala Ile Gly Gly Phe Ala Thr Leu Glu Thr Glu Ser Gly Glu Leu Val
565 570 575

ccc aca gag ctc aat gac atc tgg agg tat aac gag gag gag aag aaa 1776
Pro Thr Glu Leu Asn Asp Ile Trp Arg Tyr Asn Glu Glu Glu Lys Lys
580 585 590

tgg gag ggt gtc ctg cgg gag atc gcc tat gca gca ggt gcc acc ttc 1824
Trp Glu Gly Val Leu Arg Glu Ile Ala Tyr Ala Ala Gly Ala Thr Phe
595 600 605

cta cca gtg cgg ctc aat gtg ctg cgc ctg act aag atg tga ccagctc 1873
Leu Pro Val Arg Leu Asn Val Leu Arg Leu Thr Lys Met
610 615 620

aggcagactg aactaagcac ccctcccatc ctgcgaccct cactggcctg gccttgtggg 1933

ggctccagaa aagaggctag gagaggccag agtctacctg gatccagtta tggtgcctca 1993

ggggctgcgt cagccaagga aagggaagtg ctgcttagtc ctggactttt gggccagggt 2053

gagaaactag aggcttctcc agtgttgcca tatcccccta ggttgtcttg atccatgaac 2113

cagaaccaca g 2124

<210> 9
<211> 4414
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (991)..(2400)

<400> 9
gtcacacatg gacagcctca tctattttat tgacccattt ccagatgctg ggattaccga 60

gttacattaa tacacagctt tttatggtat acagtaatgt aatatattag cctaatgtgt 120

taggactcct taaacaaggg atactgtaat cattaataca gttgatccat taggcttggc 180

acagtgttta ttttatggaa acactctaaa agaaaaacag acctgaagtt catcatgtgt 240

aaggttacac aaatctatct ttcaaatgag ttcattactc taaataattg tcagaagtta 300

catttttgtt taatataatc tttgcagtaa ccttccaaat tagggcaatt tttctgtcca 360

gatctgccta ctgaactaga atatttaatg agcaacaact tctgttgtga attaacggac 420

tgatattggg gtttgataaa tcagtttcct aaggtttgga taagccaaga gttagactct 480

aagaatataa aaaggcctag ataacatcat aaaagtaagg ttctagatcc acaaagcaga 540

gaactagacc cagattataa gtcactccag aggtcccatt tttctggatt tgagggagga 600

agtgtttgat gacagtattt gaagtgattt tatatccatt ttcttttaaa aagtaatagt 660

ttggaattag ttacttgtta attaatctgt aaaatatttg ctgtggtgtt ataatgtgaa 720

aatgtggttg ctaattttaa aataaatttt gggtttttta aagtaaattt ttaaaaacac 780

tttttatgta tatcacagcc tcaacaagaa cttgatttaa actctccacg aaataccact 840

ttggaaagac agactttcta ttgtgttccg gtgcctgggg aatctacgtg ggtaaaagaa 900

gcctatgtta atgcaaacca agctcgagtc agtccctcaa catcctacac tcctagtcgc 960

cacaagagga gttatgaaga tgatgacgat atg gac cta cag ccc aat aag 1011
Met Asp Leu Gln Pro Asn Lys
1 5

cag aaa gac caa cat gca ggt gcc aga caa gca ggg agt gtt ggt ggt 1059
Gln Lys Asp Gln His Ala Gly Ala Arg Gln Ala Gly Ser Val Gly Gly
10 15 20

ctt caa tgg tgt gga gag cca aaa cgt tta gaa act gaa gct tct act 1107
Leu Gln Trp Cys Gly Glu Pro Lys Arg Leu Glu Thr Glu Ala Ser Thr
25 30 35

ggg caa cag ctg aac tct ctg aac ttg tct tct cct ttt gat ttg aat 1155
Gly Gln Gln Leu Asn Ser Leu Asn Leu Ser Ser Pro Phe Asp Leu Asn
40 45 50 55

ttt cca ttg cca gga gag aag ggc cct gca tgc ctt gtg aag gtt tat 1203
Phe Pro Leu Pro Gly Glu Lys Gly Pro Ala Cys Leu Val Lys Val Tyr
60 65 70

gaa gat tgg gat tgt ttc aaa gta aat gac att ctt gag cta tat ggc 1251
Glu Asp Trp Asp Cys Phe Lys Val Asn Asp Ile Leu Glu Leu Tyr Gly
75 80 85

ata ctg tct gtg gat cct gtg ctg agt ata ctg aat aat gat gaa agg 1299
Ile Leu Ser Val Asp Pro Val Leu Ser Ile Leu Asn Asn Asp Glu Arg
90 95 100

gat gcc tct gca ctg ctg gat ccg atg gag tgc aca gac aca gca gag 1347

Asp Ala Ser Ala Leu Leu Asp Pro Met Glu Cys Thr Asp Thr Ala
Glu
105 110 115

gag cag aga gta cac agt cct cct gct tca tta gtg ccg aga att
cat 1395
Glu Gln Arg Val His Ser Pro Pro Ala Ser Leu Val Pro Arg Ile
His
120 125 130
135

gtg atc tta gcc cag aag ttg caa cac atc aac cca tta ttg cct
gcc 1443
Val Ile Leu Ala Gln Lys Leu Gln His Ile Asn Pro Leu Leu Pro
Ala
140 145 150

tgc ctt aac aaa gag gag agc aaa acc tgt aag ttt gtt tca agt
ttc 1491
Cys Leu Asn Lys Glu Glu Ser Lys Thr Cys Lys Phe Val Ser Ser
Phe
155 160 165

atg tcc gaa ttg tct cca gtc aga gca gaa ctt ctt ggg ttc ctt
act 1539
Met Ser Glu Leu Ser Pro Val Arg Ala Glu Leu Leu Gly Phe Leu
Thr
170 175 180

cat gcc ctt ctg ggg gat agt ttg gct gct gaa tac ctt ata tta
cat 1587
His Ala Leu Leu Gly Asp Ser Leu Ala Ala Glu Tyr Leu Ile Leu
His
185 190 195

ctc atc tcc aca gta tat aca aga aga gat gtc ctt cca cta gga
aaa 1635
Leu Ile Ser Thr Val Tyr Thr Arg Arg Asp Val Leu Pro Leu Gly
Lys
200 205 210
215

ttt aca gtt aac ttg agt ggt tgc cca cgg aat agt acc ttc aca
gaa 1683
Phe Thr Val Asn Leu Ser Gly Cys Pro Arg Asn Ser Thr Phe Thr
Glu
220 225 230

cac ttg tat cga att att caa cat ctt gtt cca gca tct ttt cgt
ctg 1731
His Leu Tyr Arg Ile Ile Gln His Leu Val Pro Ala Ser Phe Arg
Leu

235 240 245

cag atg act ata gag aac atg aac cat ttg aaa ttc att ccc cac aaa 1779
Gln Met Thr Ile Glu Asn Met Asn His Leu Lys Phe Ile Pro His Lys
250 255 260

gac tac aca gcc aat cgc ttg gtc agt ggg ctc ctc cag ctg ccc agc 1827
Asp Tyr Thr Ala Asn Arg Leu Val Ser Gly Leu Leu Gln Leu Pro Ser
265 270 275

aat act tcc ctt gta atc gat gag act ctc ctg gaa cag ggg cag ctg 1875
Asn Thr Ser Leu Val Ile Asp Glu Thr Leu Leu Glu Gln Gly Gln Leu
280 285 290
295

gat acc cca ggt gtt cat aat gtg aca gcc ctg agc aac ctc ata acg 1923
Asp Thr Pro Gly Val His Asn Val Thr Ala Leu Ser Asn Leu Ile Thr
300 305 310

tgg cag aag gtg gat tat gac ttc agc tac cat cag atg gaa ttc ccc 1971
Trp Gln Lys Val Asp Tyr Asp Phe Ser Tyr His Gln Met Glu Phe Pro
315 320 325

tgc aat att aac gtt ttc att act tcg gag ggg agg tca ctc ctc ccg 2019
Cys Asn Ile Asn Val Phe Ile Thr Ser Glu Gly Arg Ser Leu Leu Pro
330 335 340

gca gac tgc cag att cac tta cag ccc cag cta att cca cca aac atg 2067
Ala Asp Cys Gln Ile His Leu Gln Pro Gln Leu Ile Pro Pro Asn Met
345 350 355

gag gag tac atg aac agc ctt ctc tca gcg gtg ctg cct tcc gtg ctg 2115
Glu Glu Tyr Met Asn Ser Leu Leu Ser Ala Val Leu Pro Ser Val Leu
360 365 370
375

aac aaa ttc cgc att tat cta act ctt ttg aga ttc ttg gaa tat agc 2163
Asn Lys Phe Arg Ile Tyr Leu Thr Leu Leu Arg Phe Leu Glu Tyr Ser
380 385 390

ata tct gat gaa ata acc aag gca gtt gaa gat gac ttt gtg gaa atg 2211
Ile Ser Asp Glu Ile Thr Lys Ala Val Glu Asp Asp Phe Val Glu Met
395 400 405

cgg aag aac gac cct cag agc atc act gct gat gat ctt cac cag ctg 2259
Arg Lys Asn Asp Pro Gln Ser Ile Thr Ala Asp Asp Leu His Gln Leu
410 415 420

ctc gtg gtg gct cgg tgt ctg tct ctc agt gct ggt cag aca acg ctg 2307
Leu Val Val Ala Arg Cys Leu Ser Leu Ser Ala Gly Gln Thr Thr Leu
425 430 435

tca aga gaa cga tgg ctg aga gca aag cag cta gag tct tta aga aga 2355
Ser Arg Glu Arg Trp Leu Arg Ala Lys Gln Leu Glu Ser Leu Arg Arg
440 445 450 455

acg agg ctt cag cag caa aaa tgt gtg aat gga aat gaa ctt taa aga 2403
Thr Arg Leu Gln Gln Gln Lys Cys Val Asn Gly Asn Glu Leu
460 465

tgtaatacct atgaagagta atgggcaaac tgtagccaca taattgtaaa attcagatat 2463

tcatttatac cacattgttt tataggtaat ttctatcaca aaccagtgac atttcctgaa 2523

atcaagcctg gtaacacctg atgtttatat gatattcagt aaggactttt accttactga 2583

tttcatggag cttttgaagt ttgttttata ataattatat aaattagtaa tgatgtaaaa 2643

aaagtatttg atattaaaag tttaatattg ataatgttgc tgattgtacc atttccttag 2703

cttcagctga gtcataggcc agactgttga aatgctgaaa tgaagaaggt tgttgcagtt 2763

tcaaagtcag aggaatcgtg cttcggattt cttatgtttt ctagttctct gtttttccag 2823

ttcacagtgg gttggggtgc attcagtagt ccatctttgg ggaacggagg cgtacttgcc 2883

attgattcac atgactacat gaaattctgt actgtcattt cccagatgtt tggccacaga 2943

aactttttcc cacttaacat ttgttaacag cctgcaaaac taaacttgta catggcagtg 3003

gttcccagac ttttgtattt tatggaccgg tagtaatatt tccaaaaatc tggggtacta 3063

taaggttgcc aatttacctt gccaagtaat ccgaataaat cactgtatta tcaccatttt 3123

tttcataaaa ggaaaggaca atctatctct gaataagagg agtcctttaa acggaatgaa 3183

tgtggctttt gggggcaaaa gaaaccaaga cactacattg tctttatttt ctcctatccc 3243

agtgcatttg agaaccatgc ataagggaat gctgtgctac aaagctgtgc ccaaatatga 3303

aaacaaaata ggaaacttaa aaagcaatac cccctttaga aagtttttat tttcttaaat 3363

gtcattgagt tgctttgatt ctattggatt tttggcattt tttatgggat catcagttgg 3423

ttccaagtat gttagatcag ctaacatctg ctactccagt aacagcctcg tacaactgca 3483

ggtaggtttt ctccagacca attagtttta atagagcaaa ctaacaacag actgtagtag 3543

catggttatg gcaaccagaa tcttcagaaa ggttaggaca ttacttttta agctgtcagt 3603

ggtatcaagt aacttaccta gttggaggca gataaaggat cccttacgtt ttttcctata 3663

aggcctaaat tgaaattgtt aaccaaggaa acagggtcag ccttgaaaaa tcaaggaatt 3723

cattgtacct aataactgaa gtaaaaataa ctagttgttc aacttttcct aaactcaaat 3783

ctatttttat aaacaaatgt aaataatgtt tatattagag ttgaactggt tttcattttt 3843

ataactggta gactagacct tccttaaact tttagaaata aaatgaaggc ttcactggat 3903

ttgtgaggat aaaatacatt ttctttaatt gtcctagagc aaagtacatt agtcaccatg 3963

tgtttttgt gccaatgtaa attgtaattt accaaagaaa aatacataca ttgcttggtc 4023

ttgcagaaaa gttcccttga aagaaccttt ccaataaata aaacgtccca aattagcagt 4083

accttgggct gtttttcatg agtaagaaga ttcaccatcc catgtgatct gtgtggaaaa 4143

agaccatgtc ctcttggtgg aagacatgag agagctgaac tgaagtggag gaggtggtgc 4203

aagagggacc ttcctgctca aggcccgccc aggcagcgga atagagtgca gtgcttggct 4263

gcagaaaccc tttgtccctc acctatatat acacggacag tcaagtttgt tgctctaacg 4323

taaggcacag cgttaatcct gtatggccag gaaactgagt agactcctgt gtaaccctgt 4383

ttggaacttt gccttcttaa aatgattttt c 4414

<210> 10
<211> 1467
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (46)..(1467)

<400> 10
gcctgccgta ccggtccgga attcccgggt cgacgatttc gtgcc atg tgc tcc 54

Met Cys
Ser
1

ggg ctc ctg gag ctc ctg ctg ccc atc tgg ctc tcc tgg acc ctg ggg 102
Gly Leu Leu Glu Leu Leu Leu Pro Ile Trp Leu Ser Trp Thr Leu Gly
5 10 15

acc cga ggc tct gag ccc cgc agt gtg aac gat ccc ggg aac atg tcc 150
Thr Arg Gly Ser Glu Pro Arg Ser Val Asn Asp Pro Gly Asn Met Ser
20 25 30 35

ttt gtg aag gag acg gtg gac aag ctg ttg aaa ggc tac gac att cgc 198
Phe Val Lys Glu Thr Val Asp Lys Leu Leu Lys Gly Tyr Asp Ile Arg
40 45 50

cta aga ccc gac ttc ggg ggt ccc ccg gtc tgc gtg ggg atg aac atc 246
Leu Arg Pro Asp Phe Gly Gly Pro Pro Val Cys Val Gly Met Asn Ile
55 60 65

gac atc gcc agc atc gac atg gtt tcc gaa gtc aac atg gat tat acc 294
Asp Ile Ala Ser Ile Asp Met Val Ser Glu Val Asn Met Asp Tyr Thr
70 75 80

tta acc atg tat ttt caa caa tat tgg aga gat aaa agg ctc gcc tat 342
Leu Thr Met Tyr Phe Gln Gln Tyr Trp Arg Asp Lys Arg Leu Ala Tyr
85 90 95

tct ggg atc cct ctc aac ctc acg ctt gac aat cga gtg gct gac cag 390
Ser Gly Ile Pro Leu Asn Leu Thr Leu Asp Asn Arg Val Ala Asp Gln
100 105 110 115

cta tgg gtg ccc gac aca tat ttc tta aat gac aaa aag tca ttt gtg 438
Leu Trp Val Pro Asp Thr Tyr Phe Leu Asn Asp Lys Lys Ser Phe Val

120 125 130

cat gga gtg aca gtg aaa aac cgc atg atc cgt ctt cac cct gat ggg 486
His Gly Val Thr Val Lys Asn Arg Met Ile Arg Leu His Pro Asp Gly
135 140 145

aca gtg ctg tat ggg ctc aga atc acc acg aca gca gca tgc atg atg 534
Thr Val Leu Tyr Gly Leu Arg Ile Thr Thr Thr Ala Ala Cys Met Met
150 155 160

gac ctc agg aga tac ccc ctg gac gag cag aac tgc act ctg gaa att 582
Asp Leu Arg Arg Tyr Pro Leu Asp Glu Gln Asn Cys Thr Leu Glu Ile
165 170 175

gaa agc tat ggc tac acc acg gat gac att gag ttt tac tgg cga ggc 630
Glu Ser Tyr Gly Tyr Thr Thr Asp Asp Ile Glu Phe Tyr Trp Arg Gly
180 185 190 195

ggg gac aag gct gtt acc gga gtg gaa agg att gag ctc ccg cag ttc 678
Gly Asp Lys Ala Val Thr Gly Val Glu Arg Ile Glu Leu Pro Gln Phe
200 205 210

tcc atc gtg gag cac cgt ctg gtc tcg agg aat gtt gtc ttc gcc aca 726
Ser Ile Val Glu His Arg Leu Val Ser Arg Asn Val Val Phe Ala Thr
215 220 225

ggt gcc tat cct cga ctg tca ctg agc ttt cgg ttg aag agg aac att 774
Gly Ala Tyr Pro Arg Leu Ser Leu Ser Phe Arg Leu Lys Arg Asn Ile
230 235 240

gga tac ttc att ctt cag act tat atg ccc tct ata ctg ata acg att 822
Gly Tyr Phe Ile Leu Gln Thr Tyr Met Pro Ser Ile Leu Ile Thr Ile
245 250 255

ctg tcg tgg gtg tcc ttc tgg atc aat tat gat gca tct gct gct aga 870
Leu Ser Trp Val Ser Phe Trp Ile Asn Tyr Asp Ala Ser Ala Ala Arg
260 265 270 275

gtt gcc ctc ggg atc aca act gtg ctg aca atg aca acc atc aac acc 918
Val Ala Leu Gly Ile Thr Thr Val Leu Thr Met Thr Thr Ile Asn Thr
280 285 290

cac ctt cgg gag acc ttg ccc aaa atc ccc tat gtc aaa gcc att gac 966
His Leu Arg Glu Thr Leu Pro Lys Ile Pro Tyr Val Lys Ala Ile Asp
295 300 305

atg tac ctt atg ggc tgc ttc gtc ttt gtg ttc ctg gcc ctt ctg gag 1014
Met Tyr Leu Met Gly Cys Phe Val Phe Val Phe Leu Ala Leu Leu Glu
310 315 320

tat gcc ttt gtc aac tac att ttc ttt gga aga ggc cct caa agg cag 1062
Tyr Ala Phe Val Asn Tyr Ile Phe Phe Gly Arg Gly Pro Gln Arg Gln
325 330 335

aag aag ctt gca gaa aag aca gcc aag gca aag aat gac cgt tca aag 1110
Lys Lys Leu Ala Glu Lys Thr Ala Lys Ala Lys Asn Asp Arg Ser Lys
340 345 350 355

agc gaa agc aac cgg gtg gat gct cat gga aat att ctg ttg aca tcg 1158
Ser Glu Ser Asn Arg Val Asp Ala His Gly Asn Ile Leu Leu Thr Ser
360 365 370

ctg gaa gtt cac aat gaa atg aat gag gtc tca ggc ggc att ggc gat 1206
Leu Glu Val His Asn Glu Met Asn Glu Val Ser Gly Gly Ile Gly Asp
375 380 385

acc agg aat tca gca ata tcc ttt gac aac tca gga atc cag tac agg 1254

Thr Arg Asn Ser Ala Ile Ser Phe Asp Asn Ser Gly Ile Gln Tyr Arg
390 395 400

aaa cag agc atg cct cga gaa ggg cat ggg cga ttc ctg ggg gac aga 1302
Lys Gln Ser Met Pro Arg Glu Gly His Gly Arg Phe Leu Gly Asp Arg
405 410 415

agc ctc ccg cac aag aag acc cat cta cgg agg agg tct tca cag ctc 1350
Ser Leu Pro His Lys Lys Thr His Leu Arg Arg Arg Ser Ser Gln Leu
420 425 430 435

aaa att aaa ata cct gat cta acc gat gtg aat gcc ata gac aga tgg 1398
Lys Ile Lys Ile Pro Asp Leu Thr Asp Val Asn Ala Ile Asp Arg Trp
440 445 450

tcc agg atc gtg ttt cca ttc act ttt tct ctt ttc aac tta gtt tac 1446
Ser Arg Ile Val Phe Pro Phe Thr Phe Ser Leu Phe Asn Leu Val Tyr
455 460 465

tgg ctg tac tat gtt aac tga 1467
Trp Leu Tyr Tyr Val Asn
470

<210> 11
<211> 1024
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (216)..(686)

<400> 11
gccaagatag atcaactctc cctaaaggct gacagtgaac tcttggggcc gttttattct 60

ctgaggttag caaggagtca tctactagcc attcaggagg ccagctggga
agacaaaata 120

ggcaccccaa actcagcaac ttcataacac cttcctctcc ccgcctgaag
ccttaaactg 180

catcaagtca aagaaacctg ggcaaatcc ttaac atg ttt ttg act gca
gta 233
Met Phe Leu Thr Ala
Val
1 5

aat cca cag cca ctc tct act ccg agc tgg cag att gag acc aag
tat 281
Asn Pro Gln Pro Leu Ser Thr Pro Ser Trp Gln Ile Glu Thr Lys
Tyr
10 15 20

tca acg aaa gtg ctc act gga aat tgg atg gaa gag agg aga aag
ttc 329
Ser Thr Lys Val Leu Thr Gly Asn Trp Met Glu Glu Arg Arg Lys
Phe
25 30 35

acc aga gac act gac aag aca ccc caa tcc att tac aga aaa gaa
tac 377
Thr Arg Asp Thr Asp Lys Thr Pro Gln Ser Ile Tyr Arg Lys Glu
Tyr
40 45 50

atc ccc ttc cca gac cac aga cca gac cag atc tcc agg tgg tat
ggg 425
Ile Pro Phe Pro Asp His Arg Pro Asp Gln Ile Ser Arg Trp Tyr
Gly
55 60 65
70

aag agg aaa gtt gag ggg cta cct tac aaa cac ctg atc acc ccc
cac 473
Lys Arg Lys Val Glu Gly Leu Pro Tyr Lys His Leu Ile Thr Pro
His
75 80 85

cag gag ccc cca cat cgt tac ttg atc agc acc tat gac gac cat
tac 521
Gln Glu Pro Pro His Arg Tyr Leu Ile Ser Thr Tyr Asp Asp His
Tyr
90 95 100

aac cgg cat ggt tac aac ccg ggc tgc ctc cac tcc gca ctt gga
atg 569

Asn Arg His Gly Tyr Asn Pro Gly Cys Leu His Ser Ala Leu Gly
Met
105 110 115

gac aga agt tgc tgt ggc tgc cag aga agt ctg act ttc ccc ttc
ttg 617
Asp Arg Ser Cys Cys Gly Cys Gln Arg Ser Leu Thr Phe Pro Phe
Leu
120 125 130

ctc ccc cta caa act atg gac tct atg agc agc tca agc aga gac
agc 665
Leu Pro Leu Gln Thr Met Asp Ser Met Ser Ser Ser Ser Arg Asp
Ser
135 140 145
150

tca cac cca agg ctg gcc tga ag cagagcactt atacttcatc
ctaccccaga 718
Ser His Pro Arg Leu Ala
155

ccaccgttgt gcgctatgtc ctggagggag catgcggtcc cggtccctcc
ccatcgcctg 778

catcctctcc cacacttctg agagctgcca ccccaggagc agctcagata
gaatcagctg 838

gagaccacag catcactgga cttgccagac aacaagtggc gcagataaac
tcagagtacg 898

agatctggcc cgtcaaaggt gctctcagaa tcatcatctg catttggcgg
tacctgtccc 958

ccctcaaaac ccacaggttc ctttcttttc catccaacaa ttaaagatct
ttgacactaa 1018

aaaaaa
1024

<210> 12
<211> 2556
<212> PRT
<213> Homo sapiens

<400> 12
Met Pro Pro Leu Leu Ala Pro Leu Leu Cys Leu Ala Leu Leu Pro
Ala
1 5 10 15

Leu Ala Ala Arg Gly Pro Arg Cys Ser Gln Pro Gly Glu Thr Cys
Leu
20 25 30
Asn Gly Gly Lys Cys Glu Ala Ala Asn Gly Thr Glu Ala Cys Val
Cys
35 40 45
Gly Gly Ala Phe Val Gly Pro Arg Cys Gln Asp Pro Asn Pro Cys
Leu
50 55 60
Ser Thr Pro Cys Lys Asn Ala Gly Thr Cys His Val Val Asp Arg
Arg
65 70 75
80
Gly Val Ala Asp Tyr Ala Cys Ser Cys Ala Leu Gly Phe Ser Gly
Pro
85 90 95
Leu Cys Leu Thr Pro Leu Asp Asn Ala Cys Leu Thr Asn Pro Cys
Arg
100 105 110
Asn Gly Gly Thr Cys Asp Leu Leu Thr Leu Thr Glu Tyr Lys Cys
Arg
115 120 125
Cys Pro Pro Gly Trp Ser Gly Lys Ser Cys Gln Gln Ala Asp Pro
Cys
130 135 140
Ala Ser Asn Pro Cys Ala Asn Gly Gly Gln Cys Leu Pro Phe Glu
Ala
145 150 155
160
Ser Tyr Ile Cys His Cys Pro Pro Ser Phe His Gly Pro Thr Cys
Arg
165 170 175
Gln Asp Val Asn Glu Cys Gly Gln Lys Pro Arg Leu Cys Arg His
Gly
180 185 190
Gly Thr Cys His Asn Glu Val Gly Ser Tyr Arg Cys Val Cys Arg
Ala
195 200 205
Thr His Thr Gly Pro Asn Cys Glu Arg Pro Tyr Val Pro Cys Ser
Pro
210 215 220
Ser Pro Cys Gln Asn Gly Gly Thr Cys Arg Pro Thr Gly Asp Val
Thr
225 230 235
240
His Glu Cys Ala Cys Leu Pro Gly Phe Thr Gly Gln Asn Cys Glu
Glu
245 250 255
Asn Ile Asp Asp Cys Pro Gly Asn Asn Cys Lys Asn Gly Gly Ala
Cys
260 265 270

Val Asp Gly Val Asn Thr Tyr Asn Cys Pro Cys Pro Pro Glu Trp Thr
275 280 285
Gly Gln Tyr Cys Thr Glu Asp Val Asp Glu Cys Gln Leu Met Pro Asn
290 295 300
Ala Cys Gln Asn Gly Gly Thr Cys His Asn Thr His Gly Gly Tyr Asn
305 310 315 320
Cys Val Cys Val Asn Gly Trp Thr Gly Glu Asp Cys Ser Glu Asn Ile
325 330 335
Asp Asp Cys Ala Ser Ala Ala Cys Phe His Gly Ala Thr Cys His Asp
340 345 350
Arg Val Ala Ser Phe Tyr Cys Glu Cys Pro His Gly Arg Thr Gly Leu
355 360 365
Leu Cys His Leu Asn Asp Ala Cys Ile Ser Asn Pro Cys Asn Glu Gly
370 375 380
Ser Asn Cys Asp Thr Asn Pro Val Asn Gly Lys Ala Ile Cys Thr Cys
385 390 395 400
Pro Ser Gly Tyr Thr Gly Pro Ala Cys Ser Gln Asp Val Asp Glu Cys
405 410 415
Ser Leu Gly Ala Asn Pro Cys Glu His Ala Gly Lys Cys Ile Asn Thr
420 425 430
Leu Gly Ser Phe Glu Cys Gln Cys Leu Gln Gly Tyr Thr Gly Pro Arg
435 440 445
Cys Glu Ile Asp Val Asn Glu Cys Val Ser Asn Pro Cys Gln Asn Asp
450 455 460
Ala Thr Cys Leu Asp Gln Ile Gly Glu Phe Gln Cys Met Cys Met Pro
465 470 475 480
Gly Tyr Glu Gly Val His Cys Glu Val Asn Thr Asp Glu Cys Ala Ser
485 490 495
Ser Pro Cys Leu His Asn Gly Arg Cys Leu Asp Lys Ile Asn Glu Phe
500 505 510
Gln Cys Glu Cys Pro Thr Gly Phe Thr Gly His Leu Cys Gln Tyr Asp
515 520 525

Val Asp Glu Cys Ala Ser Thr Pro Cys Lys Asn Gly Ala Lys Cys Leu
530 535 540
Asp Gly Pro Asn Thr Tyr Thr Cys Val Cys Thr Glu Gly Tyr Thr Gly
545 550 555 560
Thr His Cys Glu Val Asp Ile Asp Glu Cys Asp Pro Asp Pro Cys His
565 570 575
Tyr Gly Ser Cys Lys Asp Gly Val Ala Thr Phe Thr Cys Leu Cys Arg
580 585 590
Pro Gly Tyr Thr Gly His His Cys Glu Thr Asn Ile Asn Glu Cys Ser
595 600 605
Ser Gln Pro Cys Arg Leu Arg Gly Thr Cys Gln Asp Pro Asp Asn Ala
610 615 620
Tyr Leu Cys Phe Cys Leu Lys Gly Thr Thr Gly Pro Asn Cys Glu Ile
625 630 635 640
Asn Leu Asp Asp Cys Ala Ser Ser Pro Cys Asp Ser Gly Thr Cys Leu
645 650 655
Asp Lys Ile Asp Gly Tyr Glu Cys Ala Cys Glu Pro Gly Tyr Thr Gly
660 665 670
Ser Met Cys Asn Ser Asn Ile Asp Glu Cys Ala Gly Asn Pro Cys His
675 680 685
Asn Gly Gly Thr Cys Glu Asp Gly Ile Asn Gly Phe Thr Cys Arg Cys
690 695 700
Pro Glu Gly Tyr His Asp Pro Thr Cys Leu Ser Glu Val Asn Glu Cys
705 710 715 720
Asn Ser Asn Pro Cys Val His Gly Ala Cys Arg Asp Ser Leu Asn Gly
725 730 735
Tyr Lys Cys Asp Cys Asp Pro Gly Trp Ser Gly Thr Asn Cys Asp Ile
740 745 750
Asn Asn Asn Glu Cys Glu Ser Asn Pro Cys Val Asn Gly Gly Thr Cys
755 760 765
Lys Asp Met Thr Ser Gly Ile Val Cys Thr Cys Arg Glu Gly Phe Ser
770 775 780

Gly Pro Asn Cys Gln Thr Asn Ile Asn Glu Cys Ala Ser Asn Pro Cys
785 790 795 800
Leu Asn Lys Gly Thr Cys Ile Asp Asp Val Ala Gly Tyr Lys Cys Asn
805 810 815
Cys Leu Leu Pro Tyr Thr Gly Ala Thr Cys Glu Val Val Leu Ala Pro
820 825 830
Cys Ala Pro Ser Pro Cys Arg Asn Gly Gly Glu Cys Arg Gln Ser Glu
835 840 845
Asp Tyr Glu Ser Phe Ser Cys Val Cys Pro Thr Ala Gly Ala Lys Gly
850 855 860
Gln Thr Cys Glu Val Asp Ile Asn Glu Cys Val Leu Ser Pro Cys Arg
865 870 875 880
His Gly Ala Ser Cys Gln Asn Thr His Gly Gly Tyr Arg Cys His Cys
885 890 895
Gln Ala Gly Tyr Ser Gly Arg Asn Cys Glu Thr Asp Ile Asp Asp Cys
900 905 910
Arg Pro Asn Pro Cys His Asn Gly Gly Ser Cys Thr Asp Gly Ile Asn
915 920 925
Thr Ala Phe Cys Asp Cys Leu Pro Gly Phe Arg Gly Thr Phe Cys Glu
930 935 940
Glu Asp Ile Asn Glu Cys Ala Ser Asp Pro Cys Arg Asn Gly Ala Asn
945 950 955 960
Cys Thr Asp Cys Val Asp Ser Tyr Thr Cys Thr Cys Pro Ala Gly Phe
965 970 975
Ser Gly Ile His Cys Glu Asn Asn Thr Pro Asp Cys Thr Glu Ser Ser
980 985 990
Cys Phe Asn Gly Gly Thr Cys Val Asp Gly Ile Asn Ser Phe Thr Cys
995 1000 1005
Leu Cys Pro Pro Gly Phe Thr Gly Ser Tyr Cys Gln His Val Val Asn
1010 1015 1020
Glu Cys Asp Ser Arg Pro Cys Leu Leu Gly Gly Thr Cys Gln Asp Gly
1025 1030 1035 1040

Arg Gly Leu His Arg Cys Thr Cys Pro Gln Gly Tyr Thr Gly Pro Asn
1045 1050 1055
Cys Gln Asn Leu Val His Trp Cys Asp Ser Ser Pro Cys Lys Asn Gly
1060 1065 1070
Gly Lys Cys Trp Gln Thr His Thr Gln Tyr Arg Cys Glu Cys Pro Ser
1075 1080 1085
Gly Trp Thr Gly Leu Tyr Cys Asp Val Pro Ser Val Ser Cys Glu Val
1090 1095 1100
Ala Ala Gln Arg Gln Gly Val Asp Val Ala Arg Leu Cys Gln His Gly
1105 1110 1115 1120
Gly Leu Cys Val Asp Ala Gly Asn Thr His His Cys Arg Cys Gln Ala
1125 1130 1135
Gly Tyr Thr Gly Ser Tyr Cys Glu Asp Leu Val Asp Glu Cys Ser Pro
1140 1145 1150
Ser Pro Cys Gln Asn Gly Ala Thr Cys Thr Asp Tyr Leu Gly Gly Tyr
1155 1160 1165
Ser Cys Lys Cys Val Ala Gly Tyr His Gly Val Asn Cys Ser Glu Glu
1170 1175 1180
Ile Asp Glu Cys Leu Ser His Pro Cys Gln Asn Gly Gly Thr Cys Leu
1185 1190 1195 1200
Asp Leu Pro Asn Thr Tyr Lys Cys Ser Cys Pro Arg Gly Thr Gln Gly
1205 1210 1215
Val His Cys Glu Ile Asn Val Asp Asp Cys Asn Pro Pro Val Asp Pro
1220 1225 1230
Val Ser Arg Ser Pro Lys Cys Phe Asn Asn Gly Thr Cys Val Asp Gln
1235 1240 1245
Val Gly Gly Tyr Ser Cys Thr Cys Pro Pro Gly Phe Val Gly Glu Arg
1250 1255 1260
Cys Glu Gly Asp Val Asn Glu Cys Leu Ser Asn Pro Cys Asp Ala Arg
1265 1270 1275 1280
Gly Thr Gln Asn Cys Val Gln Arg Val Asn Asp Phe His Cys Glu Cys
1285 1290 1295

Arg Ala Gly His Thr Gly Arg Arg Cys Glu Ser Val Ile Asn Gly Cys
1300 1305 1310
Lys Gly Lys Pro Cys Lys Asn Gly Gly Thr Cys Ala Val Ala Ser Asn
1315 1320 1325
Thr Ala Arg Gly Phe Ile Cys Lys Cys Pro Ala Gly Phe Glu Gly Ala
1330 1335 1340
Thr Cys Glu Asn Asp Ala Arg Thr Cys Gly Ser Leu Arg Cys Leu Asn
1345 1350 1355
1360
Gly Gly Thr Cys Ile Ser Gly Pro Arg Ser Pro Thr Cys Leu Cys Leu
1365 1370 1375
Gly Pro Phe Thr Gly Pro Glu Cys Gln Phe Pro Ala Ser Ser Pro Cys
1380 1385 1390
Leu Gly Gly Asn Pro Cys Tyr Asn Gln Gly Thr Cys Glu Pro Thr Ser
1395 1400 1405
Glu Ser Pro Phe Tyr Arg Cys Leu Cys Pro Ala Lys Phe Asn Gly Leu
1410 1415 1420
Leu Cys His Ile Leu Asp Tyr Ser Phe Gly Gly Gly Ala Gly Arg Asp
1425 1430 1435
1440
Ile Pro Pro Pro Leu Ile Glu Glu Ala Cys Glu Leu Pro Glu Cys Gln
1445 1450 1455
Glu Asp Ala Gly Asn Lys Val Cys Ser Leu Gln Cys Asn Asn His Ala
1460 1465 1470
Cys Gly Trp Asp Gly Gly Asp Cys Ser Leu Asn Phe Asn Asp Pro Trp
1475 1480 1485
Lys Asn Cys Thr Gln Ser Leu Gln Cys Trp Lys Tyr Phe Ser Asp Gly
1490 1495 1500
His Cys Asp Ser Gln Cys Asn Ser Ala Gly Cys Leu Phe Asp Gly Phe
1505 1510 1515
1520
Asp Cys Gln Arg Ala Glu Gly Gln Cys Asn Pro Leu Tyr Asp Gln Tyr
1525 1530 1535
Cys Lys Asp His Phe Ser Asp Gly His Cys Asp Gln Gly Cys Asn Ser
1540 1545 1550

Ala Glu Cys Glu Trp Asp Gly Leu Asp Cys Ala Glu His Val Pro Glu
1555 1560 1565
Arg Leu Ala Ala Gly Thr Leu Val Val Val Val Leu Met Pro Pro Glu
1570 1575 1580
Gln Leu Arg Asn Ser Ser Phe His Phe Leu Arg Glu Leu Ser Arg Val
1585 1590 1595 1600
Leu His Thr Asn Val Val Phe Lys Arg Asp Ala His Gly Gln Gln Met
1605 1610 1615
Ile Phe Pro Tyr Tyr Gly Arg Glu Glu Glu Leu Arg Lys His Pro Ile
1620 1625 1630
Lys Arg Ala Ala Glu Gly Trp Ala Ala Pro Asp Ala Leu Leu Gly Gln
1635 1640 1645
Val Lys Ala Ser Leu Leu Pro Gly Gly Ser Glu Gly Gly Arg Arg Arg
1650 1655 1660
Arg Glu Leu Asp Pro Met Asp Val Arg Gly Ser Ile Val Tyr Leu Glu
1665 1670 1675 1680
Ile Asp Asn Arg Gln Cys Val Gln Ala Ser Ser Gln Cys Phe Gln Ser
1685 1690 1695
Ala Thr Asp Val Ala Ala Phe Leu Gly Ala Leu Ala Ser Leu Gly Ser
1700 1705 1710
Leu Asn Ile Pro Tyr Lys Ile Glu Ala Val Gln Ser Glu Thr Val Glu
1715 1720 1725
Pro Pro Pro Pro Ala Gln Leu His Phe Met Tyr Val Ala Ala Ala Ala
1730 1735 1740
Phe Val Leu Leu Phe Phe Val Gly Cys Gly Val Leu Leu Ser Arg Lys
1745 1750 1755 1760
Arg Arg Arg Gln His Gly Gln Leu Trp Phe Pro Glu Gly Phe Lys Val
1765 1770 1775
Ser Glu Ala Ser Lys Lys Lys Arg Arg Glu Pro Leu Gly Glu Asp Ser
1780 1785 1790
Val Gly Leu Lys Pro Leu Lys Asn Ala Ser Asp Gly Ala Leu Met Asp
1795 1800 1805

Asp Asn Gln Asn Glu Trp Gly Asp Glu Asp Leu Glu Thr Lys Lys Phe
1810 1815 1820
Arg Phe Glu Glu Pro Val Val Leu Pro Asp Leu Asp Asp Gln Thr Asp
1825 1830 1835 1840
His Arg Gln Trp Thr Gln Gln His Leu Asp Ala Ala Asp Leu Arg Met
1845 1850 1855
Ser Ala Met Ala Pro Thr Pro Pro Gln Gly Glu Val Asp Ala Asp Cys
1860 1865 1870
Met Asp Val Asn Val Arg Gly Pro Asp Gly Phe Thr Pro Leu Met Ile
1875 1880 1885
Ala Ser Cys Ser Gly Gly Gly Leu Glu Thr Gly Asn Ser Glu Glu Glu
1890 1895 1900
Glu Asp Ala Pro Ala Val Ile Ser Asp Phe Ile Tyr Gln Gly Ala Ser
1905 1910 1915 1920
Leu His Asn Gln Thr Asp Arg Thr Gly Glu Thr Ala Leu His Leu Ala
1925 1930 1935
Ala Arg Tyr Ser Arg Ser Asp Ala Ala Lys Arg Leu Leu Glu Ala Ser
1940 1945 1950
Ala Asp Ala Asn Ile Gln Asp Asn Met Gly Arg Thr Pro Leu His Ala
1955 1960 1965
Ala Val Ser Ala Asp Ala Gln Gly Val Phe Gln Ile Leu Ile Arg Asn
1970 1975 1980
Arg Ala Thr Asp Leu Asp Ala Arg Met His Asp Gly Thr Thr Pro Leu
1985 1990 1995 2000
Ile Leu Ala Ala Arg Leu Ala Val Glu Gly Met Leu Glu Asp Leu Ile
2005 2010 2015
Asn Ser His Ala Asp Val Asn Ala Val Asp Asp Leu Gly Lys Ser Ala
2020 2025 2030
Leu His Trp Ala Ala Ala Val Asn Asn Val Asp Ala Ala Val Val Leu
2035 2040 2045
Leu Lys Asn Gly Ala Asn Lys Asp Met Gln Asn Asn Arg Glu Glu Thr
2050 2055 2060

Pro Leu Phe Leu Ala Ala Arg Glu Gly Ser Tyr Glu Thr Ala Lys Val
2065 2070 2075
2080
Leu Leu Asp His Phe Ala Asn Arg Asp Ile Thr Asp His Met Asp Arg
2085 2090 2095
Leu Pro Arg Asp Ile Ala Gln Glu Arg Met His His Asp Ile Val Arg
2100 2105 2110
Leu Leu Asp Glu Tyr Asn Leu Val Arg Ser Pro Gln Leu His Gly Ala
2115 2120 2125
Pro Leu Gly Gly Thr Pro Thr Leu Ser Pro Pro Leu Cys Ser Pro Asn
2130 2135 2140
Gly Tyr Leu Gly Ser Leu Lys Pro Gly Val Gln Gly Lys Lys Val Arg
2145 2150 2155
2160
Lys Pro Ser Ser Lys Gly Leu Ala Cys Gly Ser Lys Glu Ala Lys Asp
2165 2170 2175
Leu Lys Ala Arg Arg Lys Lys Ser Gln Asp Gly Lys Gly Cys Leu Leu
2180 2185 2190
Asp Ser Ser Gly Met Leu Ser Pro Val Asp Ser Leu Glu Ser Pro His
2195 2200 2205
Gly Tyr Leu Ser Asp Val Ala Ser Pro Pro Leu Leu Pro Ser Pro Phe
2210 2215 2220
Gln Gln Ser Pro Ser Val Pro Leu Asn His Leu Pro Gly Met Pro Asp
2225 2230 2235
2240
Thr His Leu Gly Ile Gly His Leu Asn Val Ala Ala Lys Pro Glu Met
2245 2250 2255
Ala Ala Leu Gly Gly Gly Gly Arg Leu Ala Phe Glu Thr Gly Pro Pro
2260 2265 2270
Arg Leu Ser His Leu Pro Val Ala Ser Gly Thr Ser Thr Val Leu Gly
2275 2280 2285
Ser Ser Ser Gly Gly Ala Leu Asn Phe Thr Val Gly Gly Ser Thr Ser
2290 2295 2300
Leu Asn Gly Gln Cys Glu Trp Leu Ser Arg Leu Gln Ser Gly Met Val
2305 2310 2315
2320

Pro Asn Gln Tyr Asn Pro Leu Arg Gly Ser Val Ala Pro Gly Pro Leu
2325 2330 2335
Ser Thr Gln Ala Pro Ser Leu Gln His Gly Met Val Gly Pro Leu His
2340 2345 2350
Ser Ser Leu Ala Ala Ser Ala Leu Ser Gln Met Met Ser Tyr Gln Gly
2355 2360 2365
Leu Pro Ser Thr Arg Leu Ala Thr Gln Pro His Leu Val Gln Thr Gln
2370 2375 2380
Gln Val Gln Pro Gln Asn Leu Gln Met Gln Gln Gln Asn Leu Gln Pro
2385 2390 2395 2400
Ala Asn Ile Gln Gln Gln Gln Ser Leu Gln Pro Pro Pro Pro Pro Pro
2405 2410 2415
Gln Pro His Leu Gly Val Ser Ser Ala Ala Ser Gly His Leu Gly Arg
2420 2425 2430
Ser Phe Leu Ser Gly Glu Pro Ser Gln Ala Asp Val Gln Pro Leu Gly
2435 2440 2445
Pro Ser Ser Leu Ala Val His Thr Ile Leu Pro Gln Glu Ser Pro Ala
2450 2455 2460
Leu Pro Thr Ser Leu Pro Ser Ser Leu Val Pro Pro Val Thr Ala Ala
2465 2470 2475 2480
Gln Phe Leu Thr Pro Pro Ser Gln His Ser Tyr Ser Ser Pro Val Asp
2485 2490 2495
Asn Thr Pro Ser His Gln Leu Gln Val Pro Glu His Pro Phe Leu Thr
2500 2505 2510
Pro Ser Pro Glu Ser Pro Asp Gln Trp Ser Ser Ser Ser Pro His Ser
2515 2520 2525
Asn Val Ser Asp Trp Ser Glu Gly Val Ser Ser Pro Pro Thr Ser Met
2530 2535 2540
Gln Ser Gln Ile Ala Arg Ile Pro Glu Ala Phe Lys
2545 2550 2555

<210> 13
<211> 657
<212> PRT

<213> Homo sapiens

<400> 13

Met His Ala Leu Gly Ile Ser Ala Ser Thr Trp Thr Thr Thr Thr
Ala
1 5 10 15
Arg Pro Leu Pro Leu Pro Leu Pro Pro Ala Ala Pro Ala Ala Leu
Ser
20 25 30
Ile His Tyr Lys Ser Met Asp Ile Val Leu Thr Val Thr Met Val
His
35 40 45
Gly Lys Glu Glu Gly Leu Ile Leu Phe Asp Pro Ile Pro Val Ser
Ser
50 55 60
Gly Phe Ser Lys Asn Gly Val Leu Val Ser Val Leu Gly Thr Thr
Thr
65 70 75
80
Met Ala Val Asp Ile Pro Ala Leu Gly Val Ser Val Thr Phe Asn
Gly
85 90 95
Gln Val Phe Gln Ala Arg Leu Pro Tyr Ser Leu Phe His Asn Asn
Thr
100 105 110
Glu Gly Gln Cys Gly Thr Cys Thr Asn Asn Gln Arg Asp Asp Cys
Leu
115 120 125
Gln Arg Asp Gly Thr Thr Ala Ala Ser Cys Lys Asp Met Ala Lys
Thr
130 135 140
Trp Leu Val Pro Asp Ser Arg Lys Asp Gly Cys Trp Ala Pro Thr
Gly
145 150 155
160
Thr Pro Pro Thr Ala Ser Pro Ala Ala Pro Val Ser Ser Thr Pro
Thr
165 170 175
Pro Thr Pro Cys Pro Pro Gln Leu Leu Cys Asp Leu Met Leu Ser
Gln
180 185 190
Val Phe Ala Glu Cys His Thr Leu Leu Pro Pro Gly Pro Phe Phe
Asn
195 200 205
Ala Cys Ile Ser Asp His Cys Arg Gly Arg Leu Glu Val Pro Cys
Gln
210 215 220
Ser Leu Glu Ala Tyr Ala Glu Leu Cys Arg Ala Arg Gly Val Cys
Ser
225 230 235
240

Asp Trp Arg Gly Ala Thr Gly Gly Leu Cys Asp Leu Thr Cys Pro
Pro
245 250 255
Thr Lys Val Tyr Lys Pro Cys Gly Pro Ile Gln Pro Ala Thr Cys
Asn
260 265 270
Ser Arg Asn Gln Ser Pro Gln Leu Glu Gly Met Ala Glu Gly Cys
Phe
275 280 285
Cys Pro Glu Asp Gln Ile Leu Phe Asn Ala His Met Gly Ile Cys
Val
290 295 300
Gln Ala Cys Pro Cys Val Gly Pro Asp Gly Phe Pro Lys Phe Pro
Gly
305 310 315
320
Glu Arg Trp Val Ser Asn Cys Gln Ser Cys Val Cys Asp Glu Gly
Ser
325 330 335
Val Ser Val Gln Cys Lys Pro Leu Pro Cys Asp Ala Gln Gly Gln
Pro
340 345 350
Pro Pro Cys Asn Arg Pro Gly Phe Val Thr Val Thr Arg Pro Arg
Ala
355 360 365
Glu Asn Pro Cys Cys Pro Glu Thr Val Cys Val Cys Asn Thr Thr
Thr
370 375 380
Cys Pro Gln Ser Leu Pro Val Cys Pro Pro Gly Gln Glu Ser Ile
Cys
385 390 395
400
Thr Gln Glu Glu Gly Asp Cys Cys Pro Thr Phe Arg Cys Arg Pro
Gln
405 410 415
Leu Cys Ser Tyr Asn Gly Thr Phe Tyr Gly Val Gly Ala Thr Phe
Pro
420 425 430
Gly Ala Leu Pro Cys His Met Cys Thr Cys Leu Ser Gly Asp Thr
Gln
435 440 445
Asp Pro Thr Val Gln Cys Gln Glu Asp Ala Cys Asn Asn Thr Thr
Cys
450 455 460
Pro Gln Gly Phe Glu Tyr Lys Arg Val Ala Gly Gln Cys Cys Gly
Glu
465 470 475
480
Cys Val Gln Thr Ala Cys Leu Thr Pro Asp Gly Gln Pro Val Gln
Leu
485 490 495

Asn Glu Thr Trp Val Asn Ser His Val Asp Asn Cys Thr Val Tyr Leu
500 505 510
Cys Glu Ala Glu Gly Gly Val His Leu Leu Thr Pro Gln Pro Ala Ser
515 520 525
Cys Pro Asp Val Ser Ser Cys Arg Gly Ser Leu Arg Lys Thr Gly Cys
530 535 540
Cys Tyr Ser Cys Glu Glu Asp Ser Cys Gln Val Arg Ile Asn Thr Thr
545 550 555 560
Ile Leu Trp His Gln Gly Cys Glu Thr Glu Val Asn Ile Thr Phe Cys
565 570 575
Glu Gly Ser Cys Pro Gly Ala Ser Lys Tyr Ser Ala Glu Ala Gln Ala
580 585 590
Met Gln His Gln Cys Thr Cys Cys Gln Glu Arg Arg Val His Glu Glu
595 600 605
Thr Val Pro Leu His Cys Pro Asn Gly Ser Ala Ile Leu His Thr Tyr
610 615 620
Thr His Val Asp Glu Cys Gly Cys Thr Pro Phe Cys Val Pro Ala Pro
625 630 635 640
Met Ala Pro Pro His Thr Arg Gly Phe Pro Ala Gln Glu Ala Thr Ala
645 650 655
Val

<210> 14
<211> 184
<212> PRT
<213> Homo sapiens

<400> 14
Met Ala Thr Ala Leu Ala Leu Arg Ser Leu Tyr Arg Ala Arg Pro Ser
1 5 10 15
Leu Arg Cys Pro Pro Val Glu Leu Pro Trp Ala Pro Arg Arg Gly His
20 25 30
Arg Leu Ser Pro Ala Asp Asp Glu Leu Tyr Gln Arg Thr Arg Ile Ser

35 40 45
Leu Leu Gln Arg Glu Ala Ala Gln Ala Met Tyr Ile Asp Ser Tyr
Asn
50 55 60
Ser Arg Gly Phe Met Ile Asn Gly Asn Arg Val Leu Gly Pro Cys
Ala
65 70 75
80
Leu Leu Pro His Ser Val Val Gln Trp Asn Val Gly Ser His Gln
Asp
85 90 95
Ile Thr Glu Asp Ser Phe Ser Leu Phe Trp Leu Leu Glu Pro Arg
Ile
100 105 110
Glu Ile Val Val Val Gly Thr Gly Asp Arg Thr Glu Arg Leu Gln
Ser
115 120 125
Gln Val Leu Gln Ala Met Arg Gln Arg Gly Ile Ala Val Glu Val
Gln
130 135 140
Asp Thr Pro Asn Ala Cys Ala Thr Phe Asn Phe Leu Cys His Glu
Gly
145 150 155
160
Arg Val Thr Gly Ala Ala Leu Ile Pro Pro Pro Gly Gly Thr Ser
Leu
165 170 175
Thr Ser Leu Gly Gln Ala Ala Gln
180

<210> 15
<211> 122
<212> PRT
<213> Homo sapiens

<400> 15
Met Pro Gly His Met Pro Arg Gly Gly Trp Cys Leu Phe Thr Gly
Pro
1 5 10 15
Ala Trp Val Gly Ser Ser His Gly Glu Gly Lys Arg Arg Gly Gly
Ser
20 25 30
Leu Ala Gln Leu Gly Lys Gln Ser Gln Trp Glu His Lys Pro Val
Leu
35 40 45
Leu His Gly Gln Gly Trp Lys Lys Glu Cys Val Lys Pro Ser Pro
Met
50 55 60

Pro Leu Phe Pro Ser Val Pro Gln Met Ser Tyr Ala Ser Thr Thr Pro
65 70 75 80
Ile Ala Leu Ser Thr Pro Arg Arg Ser Ala Ser Gln Trp Arg Ser Cys
85 90 95
Ser Leu Thr Arg Ala Cys Arg Asn Met Ile Arg Ser Ser Pro Leu Ser
100 105 110
Arg Trp Leu Pro His Phe Ser Glu Thr Ser
115 120

<210> 16
<211> 527
<212> PRT
<213> Homo sapiens

<400> 16
Met Ala Gln Thr Leu Gln Met Glu Ile Pro Asn Phe Gly Asn Ser Ile
1 5 10 15
Leu Glu Cys Leu Asn Glu Gln Arg Leu Gln Gly Leu Tyr Cys Asp Val
20 25 30
Ser Val Val Val Lys Gly His Ala Phe Lys Ala His Arg Ala Val Leu
35 40 45
Ala Ala Ser Ser Ser Tyr Phe Arg Asp Leu Phe Asn Asn Ser Arg Ser
50 55 60
Ala Val Val Glu Leu Pro Ala Ala Val Gln Pro Gln Ser Phe Gln Gln
65 70 75 80
Ile Leu Ser Phe Cys Tyr Thr Gly Arg Leu Ser Met Asn Val Gly Asp
85 90 95
Gln Phe Leu Leu Met Tyr Thr Ala Gly Phe Leu Gln Ile Gln Glu Ile
100 105 110
Met Glu Lys Gly Thr Glu Phe Phe Leu Lys Val Ser Ser Pro Ser Cys
115 120 125
Asp Ser Gln Gly Leu His Ala Glu Glu Ala Pro Ser Ser Glu Pro Gln
130 135 140
Ser Pro Val Ala Gln Thr Ser Gly Trp Pro Ala Cys Ser Thr Pro Leu

145 150 155
160
Pro Leu Val Ser Arg Val Lys Thr Glu Gln Gln Glu Ser Asp Ser
Val
165 170 175
Gln Cys Met Pro Val Ala Lys Arg Leu Trp Asp Ser Gly Gln Lys
Glu
180 185 190
Ala Gly Gly Gly Gly Asn Gly Ser Arg Lys Met Ala Lys Phe Ser
Thr
195 200 205
Pro Asp Leu Ala Ala Asn Arg Pro His Gln Pro Pro Pro Pro Gln
Gln
210 215 220
Ala Pro Val Val Ala Ala Ala Gln Pro Ala Val Ala Ala Gly Ala
Gly
225 230 235
240
Gln Pro Ala Gly Gly Val Ala Ala Ala Gly Gly Val Val Ser Gly
Pro
245 250 255
Ser Thr Ser Glu Arg Thr Ser Pro Gly Thr Ser Ser Ala Tyr Thr
Ser
260 265 270
Asp Ser Pro Gly Ser Tyr His Asn Glu Glu Asp Glu Glu Glu Asp
Gly
275 280 285
Gly Glu Glu Gly Met Asp Glu Gln Tyr Arg Gln Ile Cys Asn Met
Tyr
290 295 300
Thr Met Tyr Ser Met Met Asn Val Gly Gln Thr Ala Glu Lys Val
Glu
305 310 315
320
Ala Leu Pro Glu Gln Val Ala Pro Glu Ser Arg Asn Arg Ile Arg
Val
325 330 335
Arg Gln Asp Leu Ala Ser Leu Pro Ala Glu Leu Ile Asn Gln Ile
Gly
340 345 350
Asn Arg Cys His Pro Lys Leu Tyr Asp Glu Gly Asp Pro Ser Glu
Lys
355 360 365
Leu Glu Leu Val Thr Gly Thr Asn Val Tyr Ile Thr Arg Ala Gln
Leu
370 375 380
Met Asn Cys His Val Ser Ala Gly Thr Arg His Lys Val Leu Leu
Arg
385 390 395
400
Arg Leu Leu Ala Ser Phe Phe Asp Arg Asn Thr Leu Ala Asn Ser
Cys

405 410 415
Gly Thr Gly Ile Arg Ser Ser Thr Asn Asp Pro Arg Arg Lys Pro
Leu
420 425 430
Asp Ser Arg Val Leu His Ala Val Lys Tyr Tyr Cys Gln Asn Phe
Ala
435 440 445
Pro Asn Phe Lys Glu Ser Glu Met Asn Ala Ile Ala Ala Asp Met
Cys
450 455 460
Thr Asn Ala Arg Arg Val Val Arg Lys Ser Trp Met Pro Lys Val
Lys
465 470 475
480
Val Leu Lys Ala Glu Asp Asp Ala Tyr Thr Thr Phe Ile Ser Glu
Thr
485 490 495
Gly Lys Ile Glu Pro Asp Met Met Gly Val Glu His Gly Phe Glu
Thr
500 505 510
Ala Ser His Glu Gly Glu Ala Gly Pro Ser Ala Glu Ala Leu Gln
515 520 525

<210> 17
<211> 576
<212> PRT
<213> Homo sapiens

<400> 17
Met Leu Ala Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu
Leu
1 5 10 15
Thr Tyr Gly Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu
Glu
20 25 30
Gly Ala Leu Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr
Thr
35 40 45
Ser Thr Ser Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln
Gly
50 55 60
Phe Arg Asp Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr
Leu
65 70 75
80
Asp Lys Leu Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val
Gln
85 90 95

Pro Ile Cys Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln
100 105 110
Ile His Thr Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn
115 120 125
Cys Leu Pro Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val
130 135 140
Gly Tyr Ser Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg
145 150 155 160
Lys Glu Cys Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser
165 170 175
Leu Leu Gly Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro
180 185 190
Gly Met Cys Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp
195 200 205
Tyr Asp Asn Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln
210 215 220
Gln Ile Leu Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile
225 230 235 240
Ala Tyr Gln Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe
245 250 255
Glu His Tyr Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala
260 265 270
Met Leu Ser Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu
275 280 285
Lys Thr Tyr Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp
290 295 300
Asn Gly Gly Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly
305 310 315 320
Ser Lys Gly Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val
325 330 335
His Ser Pro Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val
340 345 350

His Ile Thr Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly
Gln
355 360 365
Ile Asp Glu Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr
Ile
370 375 380
Ser Glu Gly Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile
Asp
385 390 395
400
Pro Ile Tyr Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr
Gly
405 410 415
Ile Trp Asn Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp
Lys
420 425 430
Leu Leu Thr Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln
Ser
435 440 445
Phe Ser Asn Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr
Leu
450 455 460
Ser Thr Gly Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro
Tyr
465 470 475
480
Glu Arg Val Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys
Leu
485 490 495
Leu Arg Arg Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg
Tyr
500 505 510
Pro Pro Lys Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val
Trp
515 520 525
Gly Pro Trp Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys
Asn
530 535 540
Gln Ala Glu Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys
Gln
545 550 555
560
Gln Lys Ala Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys
Gly
565 570 575

<210> 18
<211> 336

<212> PRT
<213> Homo sapiens

<400> 18
Met Glu Pro Gln Lys Ile Met Pro Pro Ser Lys Pro His Pro Pro
Val
1 5 10 15
Val Gly Lys Val Thr His His Ser Ile Glu Leu Tyr Trp Asp Leu
Glu
20 25 30
Lys Lys Ala Lys Arg Gln Gly Pro Gln Glu Gln Trp Phe Arg Phe
Ser
35 40 45
Ile Glu Glu Glu Asp Pro Lys Met His Thr Tyr Gly Ile Ile Tyr
Thr
50 55 60
Gly Tyr Ala Thr Lys His Val Val Glu Gly Leu Glu Pro Arg Thr
Leu
65 70 75
80
Tyr Arg Phe Arg Leu Lys Val Thr Ser Pro Ser Gly Glu Cys Glu
Tyr
85 90 95
Ser Pro Leu Val Ser Val Ser Thr Thr Arg Glu Pro Ile Ser Ser
Glu
100 105 110
His Leu His Arg Ala Val Ser Val Asn Asp Glu Asp Leu Leu Val
Arg
115 120 125
Ile Leu Gln Gly Gly Arg Val Lys Val Asp Val Pro Asn Lys Phe
Gly
130 135 140
Phe Thr Ala Leu Met Val Ala Ala Gln Lys Gly Tyr Thr Arg Leu
Val
145 150 155
160
Lys Ile Leu Val Ser Asn Gly Thr Asp Val Asn Leu Lys Asn Gly
Ser
165 170 175
Gly Lys Asp Ser Leu Met Leu Ala Cys Tyr Ala Gly His Leu Asp
Val
180 185 190
Val Lys Tyr Leu Arg Arg His Gly Ala Ser Trp Gln Ala Arg Asp
Leu
195 200 205
Gly Gly Cys Thr Ala Leu His Trp Ala Ala Asp Gly Gly His Cys
Ser
210 215 220
Val Ile Glu Trp Met Ile Lys Asp Gly Cys Glu Val Asp Val Val
Asp

225 230 235
240
Thr Gly Ser Gly Trp Thr Pro Leu Met Arg Val Ser Ala Val Ser
Gly
245 250 255
Asn Gln Arg Val Ala Ser Leu Leu Ile Asp Ala Gly Ala Asn Val
Asn
260 265 270
Val Lys Asp Arg Asn Gly Lys Thr Pro Leu Met Val Ala Val Leu
Asn
275 280 285
Asn His Glu Glu Leu Val Gln Leu Leu Leu Asp Lys Gly Ala Asp
Ala
290 295 300
Ser Val Lys Asn Glu Phe Gly Lys Gly Val Leu Glu Met Ala Arg
Val
305 310 315
320
Phe Asp Arg Gln Val Gly Met Leu Phe Leu Pro Ile Lys Ala Asn
Phe
325 330 335

<210> 19
<211> 621
<212> PRT
<213> Homo sapiens

<400> 19
Met Ala Leu Gly Leu Glu Gln Ala Glu Glu Gln Arg Leu Tyr Gln
Gln
1 5 10 15
Thr Leu Leu Gln Asp Gly Leu Lys Asp Met Leu Asp His Gly Lys
Phe
20 25 30
Leu Asp Cys Val Val Arg Ala Gly Glu Arg Glu Phe Pro Cys His
Arg
35 40 45
Leu Val Leu Ala Ala Cys Ser Pro Tyr Phe Arg Ala Arg Phe Leu
Ala
50 55 60
Glu Pro Glu Arg Ala Gly Glu Leu His Leu Glu Glu Val Ser Pro
Asp
65 70 75
80
Val Val Ala Gln Val Leu His Tyr Leu Tyr Thr Ser Glu Ile Ala
Leu
85 90 95

Asp Glu Ala Ser Val Gln Asp Leu Phe Ala Ala Ala His Arg Phe Gln
100 105 110
Ile Pro Ser Ile Phe Thr Ile Cys Val Ser Phe Leu Gln Lys Arg Leu
115 120 125
Cys Leu Ser Asn Cys Leu Ala Val Phe Arg Leu Gly Leu Leu Leu Asp
130 135 140
Cys Ala Arg Leu Ala Val Ala Ala Arg Asp Phe Ile Cys Ala His Phe
145 150 155 160
Thr Leu Val Ala Arg Asp Ala Asp Phe Leu Gly Leu Ser Ala Asp Glu
165 170 175
Leu Ile Ala Ile Ile Ser Ser Asp Gly Leu Asn Val Glu Lys Glu Glu
180 185 190
Ala Val Phe Glu Ala Val Met Arg Trp Ala Gly Ser Gly Asp Ala Glu
195 200 205
Ala Gln Ala Glu Arg Gln Arg Ala Leu Pro Thr Val Phe Glu Ser Val
210 215 220
Arg Cys Arg Leu Leu Pro Arg Ala Phe Leu Glu Ser Arg Val Glu Arg
225 230 235 240
His Pro Leu Val Arg Ala Gln Pro Glu Leu Leu Arg Lys Val Gln Met
245 250 255
Val Lys Asp Ala His Glu Gly Arg Ile Thr Thr Leu Arg Lys Lys Lys
260 265 270
Lys Gly Lys Asp Gly Ala Gly Ala Lys Glu Ala Asp Lys Gly Thr Ser
275 280 285
Lys Ala Lys Ala Glu Glu Asp Glu Glu Ala Glu Arg Ile Leu Pro Gly
290 295 300
Ile Leu Asn Asp Thr Leu Arg Phe Gly Met Phe Leu Gln Asp Leu Ile
305 310 315 320
Phe Met Ile Ser Glu Glu Gly Ala Val Ala Tyr Asp Pro Ala Ala Asn
325 330 335
Glu Cys Tyr Cys Ala Ser Leu Ser Asn Gln Val Pro Lys Asn His Val
340 345 350

Ser Leu Val Thr Lys Glu Asn Gln Val Phe Val Ala Gly Gly Leu Phe
355 360 365
Tyr Asn Glu Asp Asn Lys Glu Asp Pro Met Ser Ala Tyr Phe Leu Gln
370 375 380
Phe Asp His Leu Asp Ser Glu Trp Leu Gly Met Pro Pro Leu Pro Ser
385 390 395 400
Pro Arg Cys Leu Phe Gly Leu Gly Glu Ala Leu Asn Ser Ile Tyr Val
405 410 415
Val Gly Gly Arg Glu Ile Lys Asp Gly Glu Arg Cys Leu Asp Ser Val
420 425 430
Met Cys Tyr Asp Arg Leu Ser Phe Lys Trp Gly Glu Ser Asp Pro Leu
435 440 445
Pro Tyr Val Val Tyr Gly His Thr Val Leu Ser His Met Asp Leu Val
450 455 460
Tyr Val Ile Gly Gly Lys Gly Ser Asp Arg Lys Cys Leu Asn Lys Met
465 470 475 480
Cys Val Tyr Asp Pro Lys Lys Phe Glu Trp Lys Glu Leu Ala Pro Met
485 490 495
Gln Thr Ala Arg Ser Leu Phe Gly Ala Thr Val His Asp Gly Arg Ile
500 505 510
Ile Val Ala Ala Gly Val Thr Asp Thr Gly Leu Thr Ser Ser Ala Glu
515 520 525
Val Tyr Ser Ile Thr Asp Asn Lys Trp Ala Pro Phe Glu Ala Phe Pro
530 535 540
Gln Glu Arg Ser Ser Leu Ser Leu Val Ser Leu Val Gly Thr Leu Tyr
545 550 555 560
Ala Ile Gly Gly Phe Ala Thr Leu Glu Thr Glu Ser Gly Glu Leu Val
565 570 575
Pro Thr Glu Leu Asn Asp Ile Trp Arg Tyr Asn Glu Glu Glu Lys Lys
580 585 590
Trp Glu Gly Val Leu Arg Glu Ile Ala Tyr Ala Ala Gly Ala Thr Phe
595 600 605
Leu Pro Val Arg Leu Asn Val Leu Arg Leu Thr Lys Met

610 615 620

<210> 20
<211> 469
<212> PRT
<213> Homo sapiens

<400> 20
Met Asp Leu Gln Pro Asn Lys Gln Lys Asp Gln His Ala Gly Ala Arg
1 5 10 15
Gln Ala Gly Ser Val Gly Gly Leu Gln Trp Cys Gly Glu Pro Lys Arg
20 25 30
Leu Glu Thr Glu Ala Ser Thr Gly Gln Gln Leu Asn Ser Leu Asn Leu
35 40 45
Ser Ser Pro Phe Asp Leu Asn Phe Pro Leu Pro Gly Glu Lys Gly Pro
50 55 60
Ala Cys Leu Val Lys Val Tyr Glu Asp Trp Asp Cys Phe Lys Val Asn
65 70 75 80
Asp Ile Leu Glu Leu Tyr Gly Ile Leu Ser Val Asp Pro Val Leu Ser
85 90 95
Ile Leu Asn Asn Asp Glu Arg Asp Ala Ser Ala Leu Leu Asp Pro Met
100 105 110
Glu Cys Thr Asp Thr Ala Glu Glu Gln Arg Val His Ser Pro Pro Ala
115 120 125
Ser Leu Val Pro Arg Ile His Val Ile Leu Ala Gln Lys Leu Gln His
130 135 140
Ile Asn Pro Leu Leu Pro Ala Cys Leu Asn Lys Glu Glu Ser Lys Thr
145 150 155 160
Cys Lys Phe Val Ser Ser Phe Met Ser Glu Leu Ser Pro Val Arg Ala
165 170 175
Glu Leu Leu Gly Phe Leu Thr His Ala Leu Leu Gly Asp Ser Leu Ala
180 185 190
Ala Glu Tyr Leu Ile Leu His Leu Ile Ser Thr Val Tyr Thr Arg Arg
195 200 205

Asp Val Leu Pro Leu Gly Lys Phe Thr Val Asn Leu Ser Gly Cys
Pro
210 215 220
Arg Asn Ser Thr Phe Thr Glu His Leu Tyr Arg Ile Ile Gln His
Leu
225 230 235
240
Val Pro Ala Ser Phe Arg Leu Gln Met Thr Ile Glu Asn Met Asn
His
245 250 255
Leu Lys Phe Ile Pro His Lys Asp Tyr Thr Ala Asn Arg Leu Val
Ser
260 265 270
Gly Leu Leu Gln Leu Pro Ser Asn Thr Ser Leu Val Ile Asp Glu
Thr
275 280 285
Leu Leu Glu Gln Gly Gln Leu Asp Thr Pro Gly Val His Asn Val
Thr
290 295 300
Ala Leu Ser Asn Leu Ile Thr Trp Gln Lys Val Asp Tyr Asp Phe
Ser
305 310 315
320
Tyr His Gln Met Glu Phe Pro Cys Asn Ile Asn Val Phe Ile Thr
Ser
325 330 335
Glu Gly Arg Ser Leu Leu Pro Ala Asp Cys Gln Ile His Leu Gln
Pro
340 345 350
Gln Leu Ile Pro Pro Asn Met Glu Glu Tyr Met Asn Ser Leu Leu
Ser
355 360 365
Ala Val Leu Pro Ser Val Leu Asn Lys Phe Arg Ile Tyr Leu Thr
Leu
370 375 380
Leu Arg Phe Leu Glu Tyr Ser Ile Ser Asp Glu Ile Thr Lys Ala
Val
385 390 395
400
Glu Asp Asp Phe Val Glu Met Arg Lys Asn Asp Pro Gln Ser Ile
Thr
405 410 415
Ala Asp Asp Leu His Gln Leu Leu Val Val Ala Arg Cys Leu Ser
Leu
420 425 430
Ser Ala Gly Gln Thr Thr Leu Ser Arg Glu Arg Trp Leu Arg Ala
Lys
435 440 445
Gln Leu Glu Ser Leu Arg Arg Thr Arg Leu Gln Gln Gln Lys Cys
Val
450 455 460
Asn Gly Asn Glu Leu

465

<210> 21
<211> 473
<212> PRT
<213> Homo sapiens

<400> 21
Met Cys Ser Gly Leu Leu Glu Leu Leu Leu Pro Ile Trp Leu Ser Trp
1 5 10 15
Thr Leu Gly Thr Arg Gly Ser Glu Pro Arg Ser Val Asn Asp Pro Gly
20 25 30
Asn Met Ser Phe Val Lys Glu Thr Val Asp Lys Leu Leu Lys Gly Tyr
35 40 45
Asp Ile Arg Leu Arg Pro Asp Phe Gly Gly Pro Pro Val Cys Val Gly
50 55 60
Met Asn Ile Asp Ile Ala Ser Ile Asp Met Val Ser Glu Val Asn Met
65 70 75 80
Asp Tyr Thr Leu Thr Met Tyr Phe Gln Gln Tyr Trp Arg Asp Lys Arg
85 90 95
Leu Ala Tyr Ser Gly Ile Pro Leu Asn Leu Thr Leu Asp Asn Arg Val
100 105 110
Ala Asp Gln Leu Trp Val Pro Asp Thr Tyr Phe Leu Asn Asp Lys Lys
115 120 125
Ser Phe Val His Gly Val Thr Val Lys Asn Arg Met Ile Arg Leu His
130 135 140
Pro Asp Gly Thr Val Leu Tyr Gly Leu Arg Ile Thr Thr Thr Ala Ala
145 150 155 160
Cys Met Met Asp Leu Arg Arg Tyr Pro Leu Asp Glu Gln Asn Cys Thr
165 170 175
Leu Glu Ile Glu Ser Tyr Gly Tyr Thr Thr Asp Asp Ile Glu Phe Tyr
180 185 190
Trp Arg Gly Gly Asp Lys Ala Val Thr Gly Val Glu Arg Ile Glu Leu
195 200 205

Pro Gln Phe Ser Ile Val Glu His Arg Leu Val Ser Arg Asn Val
Val
210 215 220
Phe Ala Thr Gly Ala Tyr Pro Arg Leu Ser Leu Ser Phe Arg Leu
Lys
225 230 235
240
Arg Asn Ile Gly Tyr Phe Ile Leu Gln Thr Tyr Met Pro Ser Ile
Leu
245 250 255
Ile Thr Ile Leu Ser Trp Val Ser Phe Trp Ile Asn Tyr Asp Ala
Ser
260 265 270
Ala Ala Arg Val Ala Leu Gly Ile Thr Thr Val Leu Thr Met Thr
Thr
275 280 285
Ile Asn Thr His Leu Arg Glu Thr Leu Pro Lys Ile Pro Tyr Val
Lys
290 295 300
Ala Ile Asp Met Tyr Leu Met Gly Cys Phe Val Phe Val Phe Leu
Ala
305 310 315
320
Leu Leu Glu Tyr Ala Phe Val Asn Tyr Ile Phe Phe Gly Arg Gly
Pro
325 330 335
Gln Arg Gln Lys Lys Leu Ala Glu Lys Thr Ala Lys Ala Lys Asn
Asp
340 345 350
Arg Ser Lys Ser Glu Ser Asn Arg Val Asp Ala His Gly Asn Ile
Leu
355 360 365
Leu Thr Ser Leu Glu Val His Asn Glu Met Asn Glu Val Ser Gly
Gly
370 375 380
Ile Gly Asp Thr Arg Asn Ser Ala Ile Ser Phe Asp Asn Ser Gly
Ile
385 390 395
400
Gln Tyr Arg Lys Gln Ser Met Pro Arg Glu Gly His Gly Arg Phe
Leu
405 410 415
Gly Asp Arg Ser Leu Pro His Lys Lys Thr His Leu Arg Arg Arg
Ser
420 425 430
Ser Gln Leu Lys Ile Lys Ile Pro Asp Leu Thr Asp Val Asn Ala
Ile
435 440 445
Asp Arg Trp Ser Arg Ile Val Phe Pro Phe Thr Phe Ser Leu Phe
Asn
450 455 460
Leu Val Tyr Trp Leu Tyr Tyr Val Asn

465 470

<210> 22
<211> 156
<212> PRT
<213> Homo sapiens

<400> 22
Met Phe Leu Thr Ala Val Asn Pro Gln Pro Leu Ser Thr Pro Ser
Trp
1 5 10 15
Gln Ile Glu Thr Lys Tyr Ser Thr Lys Val Leu Thr Gly Asn Trp
Met
20 25 30
Glu Glu Arg Arg Lys Phe Thr Arg Asp Thr Asp Lys Thr Pro Gln
Ser
35 40 45
Ile Tyr Arg Lys Glu Tyr Ile Pro Phe Pro Asp His Arg Pro Asp
Gln
50 55 60
Ile Ser Arg Trp Tyr Gly Lys Arg Lys Val Glu Gly Leu Pro Tyr
Lys
65 70 75
80
His Leu Ile Thr Pro His Gln Glu Pro Pro His Arg Tyr Leu Ile
Ser
85 90 95
Thr Tyr Asp Asp His Tyr Asn Arg His Gly Tyr Asn Pro Gly Cys
Leu
100 105 110
His Ser Ala Leu Gly Met Asp Arg Ser Cys Cys Gly Cys Gln Arg
Ser
115 120 125
Leu Thr Phe Pro Phe Leu Leu Pro Leu Gln Thr Met Asp Ser Met
Ser
130 135 140
Ser Ser Ser Arg Asp Ser Ser His Pro Arg Leu Ala
145 150 155

<210> 23
<211> 2631
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(2631)

<223> n = a,t,c or g

<400> 23
gcggccgccg ccacgtacat gaagtgcagc tgcgccggcg ggggcggctc cacggtctca 60
ctctgcacgg cctcgatctt gtaggggatg ttgaggctgc ccagcgaggc gagcgctccc 120
aggaatgcgg ccacgtcggt ggcactctgg aagcactgcg aggaggcctg cacacactgc 180
cggttgtcaa tctccaggta gacgatggag ccgcggacgt ccatggggtc cagctccctc 240
cgccgccgcc caccctcgct gccaccaggg agcagcgagg ccttcacctg gcccagcagg 300
gcgtcaggtg cggcccagcc ctcggcggca cgcttgatgg ggtgcttgcg cagctcctcc 360
tcgcggccgt agtaagggga agatcatctg ctggccgctg agcgtcacgc tttgaagacc 420
acgttggtgt tgcagcacgc gggcttgaac tcccgcagga aagtggaaag gagctgttgc 480
gcagctgctc cggcggcatc agcaccacca ctcaccgcgt gccggccgcn cagcctctcg 540
ggtacatgct ccgcacagtc cagcccgtcc cactcgcact ccgcgctgtt gcagccctgg 600
tcgcagtgcc cgtcgctgaa gtggtccttg cagtactggt cgtacagggg gttgcactgg 660
ccttccgcac gctggcagtc aaagccgtcg aagaggcagc cggctgagtt gcactggctg 720
tcacagtggc cgtcactgaa gtacttccag cactgcagag actgcgtgca gttcttccag 780
gggtcattga agttgaggga gcagtcaccg ccgtcccagc cgcacgcgtg gttgttgcac 840
tgcaggctgc agaccttgtt gcccgcgtcc tcctggcact cgggcagctc gcacgcctcc 900
tcgatcagcg gcggggggat gtcgcgcccg gccccacccc cgaagctgta gtccaggatg 960
tggcacaaga gcccgttgaa tttggcgggg cacaggcaac ggtagaaggg gctctcggat 1020
gtgggctcac aggtcccctg gttgtagcag gggttgccgc ccaggcaggg gctgctggcc 1080
gggaactggc attcggggcc cgtgaagggg cccaggcaca ggcaggtggg gctgcgcggg 1140
ccggagatgc atgtgccgcc gttgaggcag cgcaggctgc cgcaggtacg agcgtcattc 1200
tcacacgtgg cgccctcgaa gcccgcaggg cacttgcaga tgaaacccgc gggcggtgtt 1260
ggaggccacg gcgcaggtgc ccccattctt gcagggcttg cctttgcagc cattgatgac 1320
ggactcgcag cggcgcccgg tgtgaccagc acggcactcg cagtggaagt cattgacgcg 1380
ctgcacgcag ttctgggtgc cacgggcgtc gcagggattg gacaggcact cgttgacatc 1440

cccctcacag cgctcaccca cgaagcccgg cgggcaggtg cagctgtagc cgcccacctg 1500
gtccacgcag gtgccgttgt taaagcactt ggggctccgg gacacggggt caacgggggg 1560
attgcagtcg tccacgttga tctcacagtg cacaccctga gtgccccgtg ggcaggagca 1620
cttgtaggtg ttggggaggt cgaggcaggt gcccccgttc tggcaggggt gggagaggca 1680
ctcgtcgatc tcctcagagc agttcacccc gtggtagccg gccacgcact tgcaggagta 1740
gccgcccagg tagtccgtgc aggtggcccc gttctggcag gggctgggtg agcactcgtc 1800
caccaggtcc tcacagtagc tgcctgtgta gcccgcctgg cagcggcagt ggtgcgtgtt 1860
gcccgcgtcc acacagagcc ctccatgctg gcacaggcgg gcaacgtcaa caccttgtcg 1920
ctgcgcagcc acctcacagg acacgctggg cacgtcgcag taaaggccgg tccagccgct 1980
ggggcactcg cagcggtact gggtgtgggt ctgccagcat ttgccgccgt tcttgcaggg 2040
cgaggagtca caccagtgca caaggttctg gcagttgggg ccagtgtagc cctgggggca 2100
ggtgcacctg tggagaccgc gaccgtcctg acaggtgccg cctagcaggc agggtcgtga 2160
gtcgcactca ttgactacgt gctggcagta gctgcccgtg aagccgggtg gacacaggca 2220
ggtgaacgag ttgatgccgt ccacgcaggt gccaccgttg aagcaggagc tctctgtgca 2280
gtcaggcgtg ttgttctcac agtggatccc gctgaagcct gcggggcagg tgcacgtgta 2340
gctgtccacg cagtccgtgc agttggcccc gttgcggcag gggtcactgg cacactcgtt 2400
gatgtcctcc tcacagaaag tgccccggaa gccgggcagg cagtcgcaga aggccgtgtt 2460
gatgccgtct gtgcaggagc cccgttgtg acacgggttg ggccggcagt cgtcgatgtc 2520
ggtctcgcag ttgcgcccac tgtagccggc ctggcagtgg cagcggtagc cgccgtgggt 2580
gttctggcag gatgcgccgt gccggcacgg gctcagaacg cactcgttga t 2631

<210> 24
<211> 3269
<212> DNA
<213> Homo sapiens

<400> 24
tttttttttt ttactcaatg ctcagggttt atttgcaaaa ctgggctggg ttgagcagag 60

ggggctttgg gatgctgacc cctctctgac ctggccgttt cgacctcagc cctgagttac 120
cctgcagcag gcatggacag gcggccgcct tgcctgtcgc agcttcctgg cagcagctgg 180
ccttgcctgg accagcctgt ctgccccagc ggccggacgt ggccagctgg aatccgtgct 240
gcgcgggcgc tccggccgtg gcagggacac ataggcccgc acatgcttag tttgttccag 300
ggcgtcccag tgacatcagg gaggggcaag gccaaggtgg gccacaggct ctcccagggt 360
aggactcctg agtggcgccc tcccgaaagg ggtcctgctg gcccctcacc cttctggggc 420
cggtttcacc cacggggtgg gcagcaggag cacggagcgc ggggtgccac acacaggccg 480
gggggctccg cagaggaggc ccaaggtttt catgatcaga caatgcacat cctggctcca 540
ggtggacaga gcatggggat ggaggcagaa cgttctcaga cagcagtggc ctcctgggcc 600
gggaagccac gggtgtgtgg gggagccatg ggcgcaggga cacagaaggg cgtgcagcca 660
cactcatcca cgtgggtgta ggtgtgcagg atggctgagc cgttaggaca gtgcaagggc 720
accgtctcct cgtggacccg cctctcctgg cagcaggtgc actggtgctg catggcctgg 780
gcctcttgct gagtacttgg acgcttccgg ggcaggagcc cctcgcagaa ggtgatgttg 840
acctcggtct cgcagccctg gtgccacagg gatggtcgtg ttgatgcgga cttggacagg 900
agtcctcctc acaggagtag cagcagccgg ttttcctgag gctccccctg cagctggaca 960
catctgggca ggatgcaggc tgtggggtca gcaaatggac tccaccctca gcctcacaga 1020
ggtacacggt gcagttgtcc acatggctgt tgacccaggt ttcattcagc tggactggct 1080
ggccatcggg cgtgaggcag gcggtctgga cgcactcccc acagcactgc ccggccactc 1140
tcttgtactc aaagccctgg ggacaggtag tattgttgca ggcatcctcc tgacattgca 1200
ccgttgggtc ctgggtgtcc ccagagaggc aggtacacat gtggcaggga agggcgcctg 1260
ggaaggttgc accaaccccg tagaaggtgc cattgtacga acacagctga ggtctgcagc 1320
ggaaggtggg acagcagtcg ccctcctcct gggtgcagat ggactcctgc cctggcgggc 1380
acacaggcag gctctggggg caggtggttg tgttgcacac gcacaccgtc tcggggcagc 1440
aggggttctc ggcccggggc ctggtcacgg ttacgaagcc gggacggttg cacggcgggg 1500
gctgaccctg ggcgtcacag ggcaggggct tgcactgcac cgacactgaa ccctcgtcac 1560
acacgcagga ctggcagttg ctgacccacc gctccccggg aaatttagga aacccatcgg 1620

gtcccacgca ggggcaggcc tgcacgcaga tgcccatgtg tgcgttgaag aggatctggt 1680
cctcagggca gaagcagccc tccgccatcc cctccagctg tgggctctgg ttcctagagt 1740
tgcaggtggc aggctgtatg gggccgcatg gcttgtacac tttggtgggt gggcaggtga 1800
ggtcgcacag gccaccggtt gcacctcgcc agtcactgca cactccccgg gcgcggcaga 1860
gctctgcgta agcctccagg ctctggcagg gcacctcaag gcggcccctg cagtggtcgc 1920
tgatgcaggc gttgaagaat gggcccgggg gcagaagggt gtggcactca gcaaagacct 1980
ggctcagcat cagatcacag agcagctgtg gtgggcatgg ggtgggggtg ggtgtgctag 2040
acaccggggc tgcggggctg gcagtggggg gtgtgccagt cggggcccaa cagccatcct 2100
ttctgctgtc ggggaccagc cacgtcttgg ccatgtcctt gcaactggcg gcagtggttc 2160
cgtcccgctg gagacagtcg tccctctggt tgttggtgca ggtgccgcac tggccctcgg 2220
tgttgttgtg gaagaggctg tagggcagcc gggcctggaa cacttggcca ttgaaggtga 2280
cgctcacgcc cagggcagga atgtccacag ccatggtggt ggtccccagc acagacacaa 2340
gcacgccgtt cttgctgaaa ccgctgctca ccggaattgg gtcaaacagg atcaggccct 2400
cctccttccc atgcaccatg gtgacagtga ggacgatatc catggacttg tagtggatgc 2460
tgagggccgc gggggcagcg ggcggcagcg gcagtggcag aggccgtgca gtagtggttg 2520
tccaggtaga ggctgagatt cccaaagcgt gcatggatct ctctcatgag gacataggtg 2580
cagttgcccc ggaaggtgta agaggtgccg tcaaaggtgg aatagtggga gccgccccac 2640
atgctgcaga tgcactcgca ctcatagtgg aagtcacagg gctggctcgg gtccgacact 2700
ttgatgggca ggtgcttgtt cacgcaggtg acgttggcca caggctttgg gtccagcagg 2760
acgacacggt tgtcacccac gcacctggcc accgtgcagt tctccagggt ccaggtctca 2820
ttcacctgcc ggagagggat ggcattgtca cagccagggg caggggaggg cgaggacagc 2880
ggggcggagg acactggcgg tggggaggtg ggacaggcgc cctggaagcg gtcaatgtca 2940
cagtgctgat tgcacactgc gtagaaatgg cagccggctc ggtcggtctt attgtagatg 3000
acttccccgg gcgagaaaaa ctgtccaaat gccctgcaga agcagggagt agaaaagtgg 3060
gagctgggga agccagtggg tctcagggtg gtgaggactg aggaggacac agtggacacg 3120
ctgaaggttg gcaggctgct ggggagcact gcaggggtgc gggtggtccc cacggtggac 3180

gagctgggaa ccgtggaggc agtggctgtg ggcatagtgg ccatggtggt
caccactttg 3240
gggggtgag ctgttcccag agtggagga
3269

<210> 25
<211> 920
<212> DNA
<213> Homo sapiens

<400> 25
tttttttttg tgagaaaagc cgttgaggtc ctctcaaatc tgcatgtgag
caagtaggtc 60
tacaacaggc aaacatactg cttcccgact tctgaaaagt caaggaatcc
taagtaatct 120
agatgaaaaa gtgactcctt ggtttcttca cattcccctc caagagccca
gtgagccatc 180
agctcctggc cagcaacaca cctgatacaa aatgggagaa gtgtataaat
tattatgttg 240
ataagcaaga taagtgccaa agggtagata agagtgaaag cattgggaag
cctggcagag 300
tgcagtcagc aggtcagttc ctggcggttc attgagcagc ttggcccaaa
gatgtaagtg 360
aagtccctcc tggtggaggg atgagagcag ctccagttac tcggccttca
tgacacagga 420
agttgaaggt ggcacaggca ttgggcgtgt cctgcacttc cacagcaatg
ccccgctgcc 480
tcatggcttg aagcacctgg gactgcagcc tctcggtccg gtctccagtc
cccaccacca 540
cgatctctat ccggggctcc agcaaccaga agagggaaaa gctgtcttcg
gtgatgtcct 600
ggtgggatcc cacgttccac tgcaccaccg agtgcgggag cagagcgcag
gggccgagca 660
cgcggtttcc gtttatcatg aagccgcggc tgttgtagct gtcgatgtac
attgcctgag 720
cggcctcgcg ttgcagcaga gagatgcgcg tccgctgata cagctcgtca
tccgccggcg 780
agagccgatg ccctcgccgc ggggcccagg gaagctcaac gggcggacag
cgcagcgagg 840
gtcgcgctcg gtacaagcta cgtagcgcga gagcggtggc catggctgac
caacgcgcgg 900
cgaagtcgtc accggcgccg
920

<210> 26
<211> 428
<212> DNA

<213> Homo sapiens

<400> 26
caaagggtta ggaggtctct gagaaatgag ggagccacct agacaggggt gagctcctta 60
tcatatttct gcatgccctc gtcagggagc aggacctcca ctgtgaagct gaccttcttg 120
gcgtggacaa agctataggt gttgtcgaag cgtaggacat ctgggggaca gatggaaaga 180
ggggcattgg gcttggcttt acacactctt ttttccaccc ctggccatgg aggaggacag 240
gcttgtgttc ccactggctt tgtttgccaa gctgggccaa gctcccacca cgcctcttcc 300
cttccccatg ggaactgcca acccaagcag gtcctgtaaa cagacaccaa cctcctcttg 360
gcatgtgccc aggcatggct ttactggcat cgagaacaat ggtcattaca tcccattgtc 420
tcacacct 428

<210> 27
<211> 379
<212> DNA
<213> Homo sapiens

<400> 27
ggcacgagcg acgagggcga cccctctgag aagctggagc tggtgacagg caccaacgtg 60
tacatcacaa gggcgcagct gatgaactgc cacgtcagcg caggcacgcg gcacaaggtc 120
ctactgcggc ggctcctggc ctccttcttt gaccggaaca cgctggccaa cagctgcggc 180
accggcatcc gctcttctac caacgatccc cgtcggaagc ccctggacag ccgcgtgctc 240
cacgctgtca agtactactg ccagaacttc gcccccaact tcaaggagag cgagatgaat 300
gccatcgcgg ccgacatgtg caccaacgcc cgccgcgtcg tgcgcaagag ctggatgccc 360
aaggtcaagg tgctcaagg 379

<210> 28
<211> 562
<212> DNA
<213> Homo sapiens

<400> 28
caattttttt ctcttttctt aaggtatcag atacacaccg gacttcaaca ttctatcata 60
agacctaccc aacccaactg tttacctctg gacaatgcca ccctacctca gaaactgaag 120
gaggttggat attcaacgca tatggtcgga aaatggcact tgggttttta cagaaaagaa 180
tgcatgccca ccagaagagg atttgatacc ttttttggtt cccttttggg aagtggggat 240
tactatacac actacaaatg tgacagtcct gggatgtgtg gctatgactt gtatgaaaac 300
gacaatgctg cctgggacta tgacaatggc atatactcca cacagatgta cactcagaga 360
gtacagcaaa tcttagcttc ccataacccc acaaagccta tatttttata tattgcctat 420
caagctgttc attcaccact gcaagctcct ggcaggtatt tcgaacacta ccgatccatt 480
atcaacataa acaggaggag atatgctgcc atgctttcct gcttagatga agcaatcaac 540
aacgtgacat tggctctaaa ag 562

<210> 29
<211> 577
<212> DNA
<213> Homo sapiens

<400> 29
cacagacgtg aatctgaaga atggaagtgg caaggacagt ctaatgctgg cgtgctatgc 60
gggacaccta gatgttgtga aatatctccg aagacatggc gcttcttggc aggctagaga 120
cctgggaggc tgtacagctc tgcactgggc tgcagatgga ggccactgca gtgtgattga 180
gtggatgata aaggatggct gtgaggtaga cgtcgtggac actggttcag gatggacccc 240
actcatgaga gtctctgcgg tgtcgggaaa tcagagggtg gcctctcttc taattgatgc 300
tggggccaat gtgaatgtga aggacagaaa tggaaagacg ccccttatga gtgtagtctc 360
cttattagaa gaaaggaaaa aaaagcagag gccaaagaag tcttgtgtct gctgatgaga 420
gcaccactca tctgcgaaac gcacgtaaaa caaagtgaac cgtgactgtt aaactaggga 480
tgggaaattc tgcatcttgg ggggctgtac atttatttat ttagttgaag attcactgat 540
cccactttga aatacccctt tttacctaaa aaaaaa 577

<210> 30
<211> 1083
<212> DNA
<213> Homo sapiens

<400> 30
attcccgggt cgacggctgc gagaagacga cagaagggga aggatggagc cggggccaag 60
gaggctgata agggcacaag caaagccaaa gcagaggagg atgaggaggc cgaacgtatc 120
cttcctggga tcctcaatga caccctgcgc ttcggcatgt tcctgcagga tctcatcttc 180
atgatcagtg aggagggcgc tgtggcctac gatccagcag ccaacgagtg ctactgtgct 240
tccctctcca gccaggtccc caagaaccac gtcagcctgg ttaccaagga gaaccaggtc 300
ttcgtggctg gaggcctctt ctacaacgaa gacaacaaag aggaccccat gagcgcatac 360
ttcctgcagt ttgaccatct ggactcagag tggctgggga tgccaccgct gccctcgccc 420
cgctgcctct ttggcctggg agaagctctc aactccatct acgtggtcgg tggcagagag 480
atcaaggacg gcgagcgctg cctggactcg gtcatgtgct acgacaggct gtcattcaaa 540
tggggtgaat cggacccgct gccttacgtg gtgtatggcc acacagtgct ctcccacatg 600
gaccttgtct acgtaattgg cggcaaaggc agtgacagga agtgcctgaa caagatgtgc 660
gtctatgacc ccaagaagtt tgagtggaag gagctggcac ccatgcagac cgcccgctca 720
ctctttgggg ccactgtcca tgatggccgc attatcgtgg cagctggggt caccgacaca 780
gggctgacca gttctgccga agtgtacagc atcacagaca acaagtgggc acccttcgag 840
gccttcccac aggagcgtag ctcactcagc ctggtcagcc tggtgggtac cctctatgcc 900
attggtggct ttgccacact ggagacggag tctggagagc tggttcccac agagctcaat 960
gacatctgga ggtataacga ggaggagaag aaatgggagg gtgtcctgcg ggagatcgcc 1020
tatgcagcag gtgccacctt cctaccagtg cggctcaatg tgctgtgcct gactaagatg 1080
tga 1083

<210> 31

<211> 419
<212> DNA
<213> Homo sapiens

<400> 31
ttaatcctga ctgggagaag aaagtaattg agtattttaa ggaaaagctg aaggaaaata 60
atgctcctaa gtgggtacca tcactgaacg aagttcccct tcattatttg aaacctaata 120
gttttgtgaa atttcgttgc atgattcagg atatgtttga ccctgagttt tacatgggag 180
tttatgaaac ggttaaccaa aacacaaaag cacatgttct tcattttgga aaatatagag 240
atgtagcaga gtgtgggcct caacaagaac ttgatttaaa ctctccacga aataccactt 300
tggaaagaca gactttctat tgtgttccgg tgcctgggga atctacgtgg gtaaaagaag 360
cctatgttaa tgcaaaccaa gctcgagtca gtccctcaac atctacactc ctagtcgcc 419

<210> 32
<211> 406
<212> DNA
<213> Homo sapiens

<400> 32
tttcgtgcgc catgtgctcc gggctcctgt agctcttgct gcccatctgg ctctcctgga 60
ccctggggac ccgaggctct gagcctcgca gtgtgaacga tcccgggaac atgtcctttg 120
tgaaggagac ggtggacaag ctgctgacag gcttccgctg cttcagagag cgggaggccg 180
cgcctaggcg ggctctcaga ggagccgctc taccgggaga gtcggaggct ggcgacccag 240
agtcacttag gtcttctgtg aatgcagact ggattcaata ttctgacctg tgggaagcgg 300
aggtcagtac cccgaggtgc gaagcgggct tttgccagga gtgctttagg acgccaggga 360
atcaggagaa ggatggccct ttcatttgtt aattgagctg aaacgg
406

<210> 33
<211> 476
<212> DNA
<213> Homo sapiens

<400> 33
ccaagataga tcaactctcc ctaaaggctg acagtgaact cttggggccg ttttattctc 60
tgaggttagc aaggagtcat ctactagcca ttcaggaggc cagctgggaa gacaaaatag 120
gcaccccaaa ctcagcaact tcataacacc ttcctctccc cgcctgaagc cttaaactgc 180
atcaagtcaa agaaacctgg ggcaaatcct taacatgttt ttgactgcag taaatccaca 240
gccactctct actccgagct ggcagattga gaccaagtat tcaacgaaag tgctcactgg 300
aaattggatg gaagagagga gaaaggggct accttacaaa cacctgatca cccaccacca 360
ggagcccсcа catcgctacc tgatcagcac ctatgacgac cattacaacc ggcatggtta 420
caacccgggg ctgcctccac tccgcacttg gaatggacag aagttgctgt ggctgc 476

<210> 34
<211> 874
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(874)
<223> Xaa = any amino acid or symbol as shown in the table 8 as set forth in Example 2

<400> 34
Ile Asn Glu Cys Val Leu Ser Pro Cys Arg His Gly Ala Ser Cys Gln
1 5 10 15
Asn Thr His Gly Gly Tyr Arg Cys His Cys Gln Ala Gly Tyr Ser Gly
20 25 30
Arg Asn Cys Glu Thr Asp Ile Asp Asp Cys Arg Pro Asn Pro Cys His
35 40 45
Asn Gly Gly Ser Cys Thr Asp Gly Ile Asn Thr Ala Phe Cys Asp Cys
50 55 60
Leu Pro Gly Phe Arg Gly Thr Phe Cys Glu Glu Asp Ile Asn Glu Cys
65 70 75 80
Ala Ser Asp Pro Cys Arg Asn Gly Ala Asn Cys Thr Asp Cys Val Asp

85 90 95
Ser Tyr Thr Cys Thr Cys Pro Ala Gly Phe Ser Gly Ile His Cys
Glu
100 105 110
Asn Asn Thr Pro Asp Cys Thr Glu Ser Ser Cys Phe Asn Gly Gly
Thr
115 120 125
Cys Val Asp Gly Ile Asn Ser Phe Thr Cys Leu Cys Pro Pro Gly
Phe
130 135 140
Thr Gly Ser Tyr Cys Gln His Val Val Asn Glu Cys Asp Ser Arg
Pro
145 150 155
160
Cys Leu Leu Gly Gly Thr Cys Gln Asp Gly Arg Gly Leu His Arg
Cys
165 170 175
Thr Cys Pro Gln Gly Tyr Thr Gly Pro Asn Cys Gln Asn Leu Val
His
180 185 190
Trp Cys Asp Ser Ser Pro Cys Lys Asn Gly Gly Lys Cys Trp Gln
Thr
195 200 205
His Thr Gln Tyr Arg Cys Glu Cys Pro Ser Gly Trp Thr Gly Leu
Tyr
210 215 220
Cys Asp Val Pro Ser Val Ser Cys Glu Val Ala Ala Gln Arg Gln
Gly
225 230 235
240
Val Asp Val Ala Arg Leu Cys Gln His Gly Gly Leu Cys Val Asp
Ala
245 250 255
Gly Asn Thr His His Cys Arg Cys Gln Ala Gly Tyr Thr Gly Ser
Tyr
260 265 270
Cys Glu Asp Leu Val Asp Glu Cys Ser Pro Ser Pro Cys Gln Asn
Gly
275 280 285
Ala Thr Cys Thr Asp Tyr Leu Gly Gly Tyr Ser Cys Lys Cys Val
Ala
290 295 300
Gly Tyr His Gly Val Asn Cys Ser Glu Glu Ile Asp Glu Cys Leu
Ser
305 310 315
320
His Pro Cys Gln Asn Gly Gly Thr Cys Leu Asp Leu Pro Asn Thr
Tyr
325 330 335
Lys Cys Ser Cys Pro Arg Gly Thr Gln Gly Val His Cys Glu Ile
Asn
340 345 350

Val Asp Asp Cys Asn Pro Pro Val Asp Pro Val Ser Arg Ser Pro
Lys
355 360 365
Cys Phe Asn Asn Gly Thr Cys Val Asp Gln Val Gly Gly Tyr Ser
Cys
370 375 380
Thr Cys Pro Pro Gly Phe Val Gly Glu Arg Cys Glu Gly Asp Val
Asn
385 390 395
400
Glu Cys Leu Ser Asn Pro Cys Asp Ala Arg Gly Thr Gln Asn Cys
Val
405 410 415
Gln Arg Val Asn Asp Phe His Cys Glu Cys Arg Ala Gly His Thr
Gly
420 425 430
Arg Arg Cys Glu Ser Val Ile Asn Gly Cys Lys Gly Lys Pro Cys
Lys
435 440 445
Asn Gly Gly Thr Cys Ala Val Ala Ser Asn Thr Ala Arg Gly Phe
Ile
450 455 460
Cys Lys Cys Pro Ala Gly Phe Glu Gly Ala Thr Cys Glu Asn Asp
Ala
465 470 475
480
Arg Thr Cys Gly Ser Leu Arg Cys Leu Asn Gly Gly Thr Cys Ile
Ser
485 490 495
Gly Pro Arg Ser Pro Thr Cys Leu Cys Leu Gly Pro Phe Thr Gly
Pro
500 505 510
Glu Cys Gln Phe Pro Ala Ser Ser Pro Cys Leu Gly Gly Asn Pro
Cys
515 520 525
Tyr Asn Gln Gly Thr Cys Glu Pro Thr Ser Glu Ser Pro Phe Tyr
Arg
530 535 540
Cys Leu Cys Pro Ala Lys Phe Asn Gly Leu Leu Cys His Ile Leu
Asp
545 550 555
560
Tyr Ser Phe Gly Gly Gly Ala Gly Arg Asp Ile Pro Pro Pro Leu
Ile
565 570 575
Glu Glu Ala Cys Glu Leu Pro Glu Cys Gln Glu Asp Ala Gly Asn
Lys
580 585 590
Val Cys Ser Leu Gln Cys Asn Asn His Ala Cys Gly Trp Asp Gly
Gly
595 600 605

Asp Cys Ser Leu Asn Phe Asn Asp Pro Trp Lys Asn Cys Thr Gln
Ser
610 615 620
Leu Gln Cys Trp Lys Tyr Phe Ser Asp Gly His Cys Asp Ser Gln
Cys
625 630 635
640
Asn Ser Ala Gly Cys Leu Phe Asp Gly Phe Asp Cys Gln Arg Ala
Glu
645 650 655
Gly Gln Cys Asn Pro Leu Tyr Asp Gln Tyr Cys Lys Asp His Phe
Ser
660 665 670
Asp Gly His Cys Asp Gln Gly Cys Asn Ser Ala Glu Cys Glu Trp
Asp
675 680 685
Gly Leu Asp Cys Ala Glu His Val Pro Glu Arg Leu Ala Ala Gly
Thr
690 695 700
Arg Xaa Val Val Val Leu Met Pro Pro Glu Gln Leu Arg Asn Ser
Ser
705 710 715
720
Phe His Phe Pro Ala Gly Val Gln Ala Arg Val Leu His Thr Asn
Val
725 730 735
Val Phe Lys Arg Asp Ala His Gly Gln Gln Met Ile Phe Pro Tyr
Tyr
740 745 750
Gly Arg Glu Glu Glu Leu Arg Lys His Pro Ile Lys Arg Ala Ala
Glu
755 760 765
Gly Trp Ala Ala Pro Asp Ala Leu Leu Gly Gln Val Lys Ala Ser
Leu
770 775 780
Leu Pro Gly Gly Ser Glu Gly Gly Arg Arg Arg Arg Glu Leu Asp
Pro
785 790 795
800
Met Asp Val Arg Gly Ser Ile Val Tyr Leu Glu Ile Asp Asn Arg
Gln
805 810 815
Cys Val Gln Ala Ser Ser Gln Cys Phe Gln Ser Ala Thr Asp Val
Ala
820 825 830
Ala Phe Leu Gly Ala Leu Ala Ser Leu Gly Ser Leu Asn Ile Pro
Tyr
835 840 845
Lys Ile Glu Ala Val Gln Ser Glu Thr Val Glu Pro Pro Pro Pro
Ala
850 855 860
Gln Leu His Phe Met Tyr Val Ala Ala Ala

865 870

<210> 35
<211> 895
<212> PRT
<213> Homo sapiens

<400> 35
Ser Ser Thr Leu Gly Thr Ala His Thr Pro Lys Val Val Thr Thr Met
1 5 10 15
Ala Thr Met Pro Thr Ala Thr Ala Ser Thr Val Pro Ser Ser Ser Thr
20 25 30
Val Gly Thr Thr Arg Thr Pro Ala Val Leu Pro Ser Ser Leu Pro Thr
35 40 45
Phe Ser Val Ser Thr Val Ser Ser Ser Val Leu Thr Thr Leu Arg Pro
50 55 60
Thr Gly Phe Pro Ser Ser His Phe Ser Thr Pro Cys Phe Cys Arg Ala
65 70 75 80
Phe Gly Gln Phe Phe Ser Pro Gly Glu Val Ile Tyr Asn Lys Thr Asp
85 90 95
Arg Ala Gly Cys His Phe Tyr Ala Val Cys Asn Gln His Cys Asp Ile
100 105 110
Asp Arg Phe Gln Gly Ala Cys Pro Thr Ser Pro Pro Pro Val Ser Ser
115 120 125
Ala Pro Leu Ser Ser Pro Ser Pro Ala Pro Gly Cys Asp Asn Ala Ile
130 135 140
Pro Leu Arg Gln Val Asn Glu Thr Trp Thr Leu Glu Asn Cys Thr Val
145 150 155 160
Ala Arg Cys Val Gly Asp Asn Arg Val Val Leu Leu Asp Pro Lys Pro
165 170 175
Val Ala Asn Val Thr Cys Val Asn Lys His Leu Pro Ile Lys Val Ser
180 185 190
Asp Pro Ser Gln Pro Cys Asp Phe His Tyr Glu Cys Glu Cys Ile Cys
195 200 205

Ser Met Trp Gly Gly Ser His Tyr Ser Thr Phe Asp Gly Thr Ser Tyr
210 215 220
Thr Phe Arg Gly Asn Cys Thr Tyr Val Leu Met Arg Glu Ile His Ala
225 230 235 240
Arg Phe Gly Asn Leu Ser Leu Tyr Leu Asp Asn His Tyr Cys Thr Ala
245 250 255
Ser Ala Thr Ala Ala Ala Ala Arg Cys Pro Ala Ala Leu Ser Ile His
260 265 270
Tyr Lys Ser Met Asp Ile Val Leu Thr Val Thr Met Val His Gly Lys
275 280 285
Glu Glu Gly Leu Ile Leu Phe Asp Pro Ile Pro Val Ser Ser Gly Phe
290 295 300
Ser Lys Asn Gly Val Leu Val Ser Val Leu Gly Thr Thr Thr Met Ala
305 310 315 320
Val Asp Ile Pro Ala Leu Gly Val Ser Val Thr Phe Asn Gly Gln Val
325 330 335
Phe Gln Ala Arg Leu Pro Tyr Ser Leu Phe His Asn Asn Thr Glu Gly
340 345 350
Gln Cys Gly Thr Cys Thr Asn Asn Gln Arg Asp Asp Cys Leu Gln Arg
355 360 365
Asp Gly Thr Thr Ala Ala Ser Cys Lys Asp Met Ala Lys Thr Trp Leu
370 375 380
Val Pro Asp Ser Arg Lys Asp Gly Cys Trp Ala Pro Thr Gly Thr Pro
385 390 395 400
Pro Thr Ala Ser Pro Ala Ala Pro Val Ser Ser Thr Pro Thr Pro Thr
405 410 415
Pro Cys Pro Pro Gln Leu Leu Cys Asp Leu Met Leu Ser Gln Val Phe
420 425 430
Ala Glu Cys His Thr Leu Leu Pro Pro Gly Pro Phe Phe Asn Ala Cys
435 440 445
Ile Ser Asp His Cys Arg Gly Arg Leu Glu Val Pro Cys Gln Ser Leu
450 455 460

Glu Ala Tyr Ala Glu Leu Cys Arg Ala Arg Gly Val Cys Ser Asp
Trp
465 470 475
480
Arg Gly Ala Thr Gly Gly Leu Cys Asp Leu Thr Cys Pro Pro Thr
Lys
485 490 495
Val Tyr Lys Pro Cys Gly Pro Ile Gln Pro Ala Thr Cys Asn Ser
Arg
500 505 510
Asn Gln Ser Pro Gln Leu Glu Gly Met Ala Glu Gly Cys Phe Cys
Pro
515 520 525
Glu Asp Gln Ile Leu Phe Asn Ala His Met Gly Ile Cys Val Gln
Ala
530 535 540
Cys Pro Cys Val Gly Pro Asp Gly Phe Pro Lys Phe Pro Gly Glu
Arg
545 550 555
560
Trp Val Ser Asn Cys Gln Ser Cys Val Cys Asp Glu Gly Ser Val
Ser
565 570 575
Val Gln Cys Lys Pro Leu Pro Cys Asp Ala Gln Gly Gln Pro Pro
Pro
580 585 590
Cys Asn Arg Pro Gly Phe Val Thr Val Thr Arg Pro Arg Ala Glu
Asn
595 600 605
Pro Cys Cys Pro Glu Thr Val Cys Val Cys Asn Thr Thr Thr Cys
Pro
610 615 620
Gln Ser Leu Pro Val Cys Pro Pro Gly Gln Glu Ser Ile Cys Thr
Gln
625 630 635
640
Glu Glu Gly Asp Cys Cys Pro Thr Phe Arg Cys Arg Pro Gln Leu
Cys
645 650 655
Ser Tyr Asn Gly Thr Phe Tyr Gly Val Gly Ala Thr Phe Pro Gly
Ala
660 665 670
Leu Pro Cys His Met Cys Thr Cys Leu Ser Gly Asp Thr Gln Asp
Pro
675 680 685
Thr Val Gln Cys Gln Glu Asp Ala Cys Asn Asn Thr Thr Cys Pro
Gln
690 695 700
Gly Phe Glu Tyr Lys Arg Val Ala Gly Gln Cys Cys Gly Glu Cys
Val
705 710 715
720

Gln Thr Ala Cys Leu Thr Pro Asp Gly Gln Pro Val Gln Leu Asn Glu
725 730 735
Thr Trp Val Asn Ser His Val Asp Asn Cys Thr Val Tyr Leu Cys Glu
740 745 750
Ala Glu Gly Gly Val His Leu Leu Thr Pro Gln Pro Ala Ser Cys Pro
755 760 765
Asp Val Ser Ser Cys Arg Gly Ser Leu Arg Lys Thr Gly Cys Cys Tyr
770 775 780
Ser Cys Glu Glu Asp Ser Cys Gln Val Arg Ile Asn Thr Thr Ile Leu
785 790 795 800
Trp His Gln Gly Cys Glu Thr Glu Val Asn Ile Thr Phe Cys Glu Gly
805 810 815
Ser Cys Pro Gly Ala Ser Lys Tyr Ser Ala Glu Ala Gln Ala Met Gln
820 825 830
His Gln Cys Thr Cys Cys Gln Glu Arg Arg Val His Glu Glu Thr Val
835 840 845
Pro Leu His Cys Pro Asn Gly Ser Ala Ile Leu His Thr Tyr Thr His
850 855 860
Val Asp Glu Cys Gly Cys Thr Pro Phe Cys Val Pro Ala Pro Met Ala
865 870 875 880
Pro Pro His Thr Arg Gly Phe Pro Ala Gln Glu Ala Thr Ala Val
885 890 895

<210> 36
<211> 196
<212> PRT
<213> Homo sapiens

<400> 36
Ala Pro Val Thr Thr Ser Pro Arg Val Gly Gln Pro Trp Arg Thr Ala
1 5 10 15
Leu Ala Leu Arg Ser Leu Tyr Arg Ala Arg Pro Ser Leu Arg Cys Pro
20 25 30
Pro Val Glu Leu Pro Trp Ala Pro Arg Arg Gly His Arg Leu Ser Pro

35 40 45
Ala Asp Asp Glu Leu Tyr Gln Arg Thr Arg Ile Ser Leu Leu Gln
Arg
50 55 60
Glu Ala Ala Gln Ala Met Tyr Ile Asp Ser Tyr Asn Ser Arg Gly
Phe
65 70 75
80
Met Ile Asn Gly Asn Arg Val Leu Gly Pro Cys Ala Leu Leu Pro
His
85 90 95
Ser Val Val Gln Trp Asn Val Gly Ser His Gln Asp Ile Thr Glu
Asp
100 105 110
Ser Phe Ser Leu Phe Trp Leu Leu Glu Pro Arg Ile Glu Ile Val
Val
115 120 125
Val Gly Thr Gly Asp Arg Thr Glu Arg Leu Gln Ser Gln Val Leu
Gln
130 135 140
Ala Met Arg Gln Arg Gly Ile Ala Val Glu Val Gln Asp Thr Pro
Asn
145 150 155
160
Ala Cys Ala Thr Phe Asn Phe Leu Cys His Glu Gly Arg Val Thr
Gly
165 170 175
Ala Ala Leu Ile Pro Pro Pro Gly Gly Thr Ser Leu Thr Ser Leu
Gly
180 185 190
Gln Ala Ala Gln
195

<210> 37
<211> 113
<212> PRT
<213> Homo sapiens

<400> 37
Ser His Ala Trp Ala His Ala Lys Arg Arg Leu Val Ser Val Tyr
Arg
1 5 10 15
Thr Cys Leu Gly Trp Gln Phe Pro Trp Gly Arg Glu Glu Ala Trp
Trp
20 25 30
Glu Leu Gly Pro Ala Trp Gln Thr Lys Pro Val Gly Thr Gln Ala
Cys
35 40 45

Pro Pro Pro Trp Pro Gly Val Glu Lys Arg Val Cys Lys Ala Lys Pro
50 55 60
Asn Ala Pro Leu Ser Ile Cys Pro Pro Asp Val Leu Arg Phe Asp Asn
65 70 75 80
Thr Tyr Ser Phe Val His Ala Lys Lys Val Ser Phe Thr Val Glu Val
85 90 95
Leu Leu Pro Asp Glu Gly Met Gln Lys Tyr Asp Lys Glu Leu Thr Pro
100 105 110
Val

<210> 38
<211> 126
<212> PRT
<213> Homo sapiens

<400> 38
Gly Thr Ser Asp Glu Gly Asp Pro Ser Glu Lys Leu Glu Leu Val Thr
1 5 10 15
Gly Thr Asn Val Tyr Ile Thr Arg Ala Gln Leu Met Asn Cys His Val
20 25 30
Ser Ala Gly Thr Arg His Lys Val Leu Leu Arg Arg Leu Leu Ala Ser
35 40 45
Phe Phe Asp Arg Asn Thr Leu Ala Asn Ser Cys Gly Thr Gly Ile Arg
50 55 60
Ser Ser Thr Asn Asp Pro Arg Arg Lys Pro Leu Asp Ser Arg Val Leu
65 70 75 80
His Ala Val Lys Tyr Tyr Cys Gln Asn Phe Ala Pro Asn Phe Lys Glu
85 90 95
Ser Glu Met Asn Ala Ile Ala Ala Asp Met Cys Thr Asn Ala Arg Arg
100 105 110
Val Val Arg Lys Ser Trp Met Pro Lys Val Lys Val Leu Lys
115 120 125

<210> 39
<211> 187
<212> PRT
<213> Homo sapiens

<400> 39
Gln Phe Phe Ser Leu Phe Leu Arg Tyr Gln Ile His Thr Gly Leu
Gln
1 5 10 15
His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro Leu Asp
Asn
20 25 30
Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser Thr His
Met
35 40 45
Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys Met Pro
Thr
50 55 60
Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly Ser Gly
Asp
65 70 75
80
Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys Gly Tyr
Asp
85 90 95
Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn Gly Ile
Tyr
100 105 110
Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu Ala Ser
His
115 120 125
Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln Ala Val
His
130 135 140
Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr Arg Ser
Ile
145 150 155
160
Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser Cys Leu
Asp
165 170 175
Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys
180 185

<210> 40
<211> 137
<212> PRT
<213> Homo sapiens

<400> 40
Thr Asp Val Asn Leu Lys Asn Gly Ser Gly Lys Asp Ser Leu Met Leu
1 5 10 15
Ala Cys Tyr Ala Gly His Leu Asp Val Val Lys Tyr Leu Arg Arg His
20 25 30
Gly Ala Ser Trp Gln Ala Arg Asp Leu Gly Gly Cys Thr Ala Leu His
35 40 45
Trp Ala Ala Asp Gly Gly His Cys Ser Val Ile Glu Trp Met Ile Lys
50 55 60
Asp Gly Cys Glu Val Asp Val Val Asp Thr Gly Ser Gly Trp Thr Pro
65 70 75 80
Leu Met Arg Val Ser Ala Val Ser Gly Asn Gln Arg Val Ala Ser Leu
85 90 95
Leu Ile Asp Ala Gly Ala Asn Val Asn Val Lys Asp Arg Asn Gly Lys
100 105 110
Thr Pro Leu Met Ser Val Val Ser Leu Leu Glu Glu Arg Lys Lys Lys
115 120 125
Gln Arg Pro Lys Lys Ser Cys Val Cys
130 135

<210> 41
<211> 360
<212> PRT
<213> Homo sapiens

<400> 41
Ile Pro Gly Ser Thr Ala Ala Arg Arg Arg Gln Lys Gly Lys Asp Gly
1 5 10 15
Ala Gly Ala Lys Glu Ala Asp Lys Gly Thr Ser Lys Ala Lys Ala Glu
20 25 30
Glu Asp Glu Glu Ala Glu Arg Ile Leu Pro Gly Ile Leu Asn Asp Thr
35 40 45
Leu Arg Phe Gly Met Phe Leu Gln Asp Leu Ile Phe Met Ile Ser Glu
50 55 60

Glu Gly Ala Val Ala Tyr Asp Pro Ala Ala Asn Glu Cys Tyr Cys Ala
65 70 75 80
Ser Leu Ser Ser Gln Val Pro Lys Asn His Val Ser Leu Val Thr Lys
85 90 95
Glu Asn Gln Val Phe Val Ala Gly Gly Leu Phe Tyr Asn Glu Asp Asn
100 105 110
Lys Glu Asp Pro Met Ser Ala Tyr Phe Leu Gln Phe Asp His Leu Asp
115 120 125
Ser Glu Trp Leu Gly Met Pro Pro Leu Pro Ser Pro Arg Cys Leu Phe
130 135 140
Gly Leu Gly Glu Ala Leu Asn Ser Ile Tyr Val Val Gly Gly Arg Glu
145 150 155 160
Ile Lys Asp Gly Glu Arg Cys Leu Asp Ser Val Met Cys Tyr Asp Arg
165 170 175
Leu Ser Phe Lys Trp Gly Glu Ser Asp Pro Leu Pro Tyr Val Val Tyr
180 185 190
Gly His Thr Val Leu Ser His Met Asp Leu Val Tyr Val Ile Gly Gly
195 200 205
Lys Gly Ser Asp Arg Lys Cys Leu Asn Lys Met Cys Val Tyr Asp Pro
210 215 220
Lys Lys Phe Glu Trp Lys Glu Leu Ala Pro Met Gln Thr Ala Arg Ser
225 230 235 240
Leu Phe Gly Ala Thr Val His Asp Gly Arg Ile Ile Val Ala Ala Gly
245 250 255
Val Thr Asp Thr Gly Leu Thr Ser Ser Ala Glu Val Tyr Ser Ile Thr
260 265 270
Asp Asn Lys Trp Ala Pro Phe Glu Ala Phe Pro Gln Glu Arg Ser Ser
275 280 285
Leu Ser Leu Val Ser Leu Val Gly Thr Leu Tyr Ala Ile Gly Gly Phe
290 295 300
Ala Thr Leu Glu Thr Glu Ser Gly Glu Leu Val Pro Thr Glu Leu Asn
305 310 315 320

Asp Ile Trp Arg Tyr Asn Glu Glu Glu Lys Lys Trp Glu Gly Val
Leu
325 330 335
Arg Glu Ile Ala Tyr Ala Ala Gly Ala Thr Phe Leu Pro Val Arg
Leu
340 345 350
Asn Val Leu Cys Leu Thr Lys Met
355 360

<210> 42
<211> 139
<212> PRT
<213> Homo sapiens

<400> 42
Asn Pro Asp Trp Glu Lys Lys Val Ile Glu Tyr Phe Lys Glu Lys
Leu
1 5 10 15
Lys Glu Asn Asn Ala Pro Lys Trp Val Pro Ser Leu Asn Glu Val
Pro
20 25 30
Leu His Tyr Leu Lys Pro Asn Ser Phe Val Lys Phe Arg Cys Met
Ile
35 40 45
Gln Asp Met Phe Asp Pro Glu Phe Tyr Met Gly Val Tyr Glu Thr
Val
50 55 60
Asn Gln Asn Thr Lys Ala His Val Leu His Phe Gly Lys Tyr Arg
Asp
65 70 75
80
Val Ala Glu Cys Gly Pro Gln Gln Glu Leu Asp Leu Asn Ser Pro
Arg
85 90 95
Asn Thr Thr Leu Glu Arg Gln Thr Phe Tyr Cys Val Pro Val Pro
Gly
100 105 110
Glu Ser Thr Trp Val Lys Glu Ala Tyr Val Asn Ala Asn Gln Ala
Arg
115 120 125
Val Ser Pro Ser Thr Ser Thr Leu Leu Val Ala
130 135

<210> 43
<211> 129
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(129)
<223> Xaa = any amino acid or symbol as shown in the table 8 as set forth in Example 2

<400> 43
Ser Cys Ala Met Cys Ser Gly Leu Leu Xaa Leu Leu Leu Pro Ile Trp
1 5 10 15
Leu Ser Trp Thr Leu Gly Thr Arg Gly Ser Glu Pro Arg Ser Val Asn
20 25 30
Asp Pro Gly Asn Met Ser Phe Val Lys Glu Thr Val Asp Lys Leu Leu
35 40 45
Thr Gly Phe Arg Cys Phe Arg Glu Arg Glu Ala Ala Pro Arg Arg Ala
50 55 60
Leu Arg Gly Ala Ala Leu Pro Gly Glu Ser Glu Ala Gly Asp Pro Glu
65 70 75 80
Ser Leu Arg Ser Ser Val Asn Ala Asp Trp Ile Gln Tyr Ser Asp Leu
85 90 95
Trp Glu Ala Glu Val Ser Thr Pro Arg Cys Glu Ala Gly Phe Cys Gln
100 105 110
Glu Cys Phe Arg Thr Pro Gly Asn Gln Glu Lys Asp Gly Pro Phe Ile
115 120 125
Cys

<210> 44
<211> 146
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(146)
<223> Xaa = any amino acid or symbol as shown in the table 8 as set forth in Example 2

<400> 44

Thr Leu Gly Ala Val Leu Phe Ser Glu Val Ser Lys Glu Ser Ser Thr
1 5 10 15
Ser His Ser Gly Gly Gln Leu Gly Arg Gln Asn Arg His Pro Lys Leu
20 25 30
Ser Asn Phe Ile Thr Pro Ser Ser Pro Arg Leu Lys Pro Xaa Thr Ala
35 40 45
Ser Ser Gln Arg Asn Leu Gly Gln Ile Leu Asn Met Phe Leu Thr Ala
50 55 60
Val Asn Pro Gln Pro Leu Ser Thr Pro Ser Trp Gln Ile Glu Thr Lys
65 70 75 80
Tyr Ser Thr Lys Val Leu Thr Gly Asn Trp Met Glu Glu Arg Arg Lys
85 90 95
Gly Leu Pro Tyr Lys His Leu Ile Thr His His Gln Glu Pro Pro His
100 105 110
Arg Tyr Leu Ile Ser Thr Tyr Asp Asp His Tyr Asn Arg His Gly Tyr
115 120 125
Asn Pro Gly Leu Pro Pro Leu Arg Thr Trp Asn Gly Gln Lys Leu Leu
130 135 140
Trp Leu
145